Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/GB05/000115

International filing date: 14 January 2005 (14.01.2005)

Document type: Certified copy of priority document

Document details: Country/Office: GB
Number: 0400814.0
Filing date: 14 January 2004 (14.01.2004)

Date of receipt at the International Bureau: 04 March 2005 (04.03.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

Dated   28 January 2005

Patents Form 1/77

Patents Act 1977
(Rule 16)

PATENTS · DESIGNS
The Patent Office
COPYRIGHT · TRADE MARKS

1/77

1ETANA4 E948422-2 002004
P01/7700 0.00-0400814.0 NONE

Request for grant of a patent
(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office

Cardiff Road
Newport
South Wales
NP10 8QQ

1. Your reference	REP07664GB	
2. Patent application number *(The Patent Office will fill this part in)*	**0400814.0**	1 4 JAN 2004
3. Full name, address and postcode of the or of each applicant *(underline all surnames)*	Ark Therapeutics Ltd 1 Fitzroy Mews London W1T 6DE	

8100612003

Patents ADP number *(if you know it)*	
If the applicant is a corporate body, give the country/state of its incorporation	GB
4. Title of the invention	Integrating Gene Therapy Vector
5. Name of your agent *(if you have one)*	Gill Jennings & Every
"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*	Broadgate House 7 Eldon Street London EC2M 7LH
Patents ADP number *(if you know it)*	745002 ✓

6. Priority: Complete this section if you are declaring priority from one or more earlier patent applications, filed in the last 12 months.	Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. Divisionals, etc: Complete this section only if this application is a divisional application or resulted from an entitlement dispute (see note f)	Number of earlier UK application	Date of filing *(day / month / year)*

8. Is a Patents Form 7/77 (Statement of inventorship and of right to grant of a patent) required in support of this request?   **YES**

Answer YES if:
a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
Otherwise answer NO (See note d)

Patents Form 1/77

9. Accompanying documents: A patent application
   must include a description of the invention.
   Not counting duplicates, please enter the number
   of pages of each item accompanying this form:

   Continuation sheets of this form

   Description   77

   Claim(s)   0

   Abstract   0

   Drawing(s)   0

10. If you are also filing any of the following,
    state how many against each item.

    Priority documents

    Translations of priority documents

    Statement of Inventorship and right
    to grant of a patent (Patents Form 7/77)

    Request for a preliminary examination
    and search (Patents Form 9/77)

    Request for a substantive examination
    (Patents Form 10/77)   NO

    Any other documents (please specify) /

11. I/We request the grant of a patent on the basis of this application.
    For the applicant
    Gill Jennings & Every
    Signature   Date 14 January 2004

12. Name, daytime telephone number and
    e-mail address, if any, of person to contact in
    the United Kingdom
        PERRY Robert Edward
        020 7377 1377

Warning
After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes
a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 08459 500505.

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you have answered YES in part 8, a Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) Part 7 should only be completed when a divisional application is being made under section 15(4), or when an application is being made under section 8(3), 12(6) or 37(4) following an entitlement dispute. By completing part 7 you are requesting that this application takes the same filing date as an earlier UK application. If you want the new application to have the same priority date(s) as the earlier UK application, you should also complete part 6 with the priority details.

1. TITLE: INTEGRATING GENE THERAPY VECTOR

2. BACKGROUND

Integrating vectors and especially retrovirus-based vectors designed for gene therapy have gained much undesirable publicity because of the side effects associated to
5   retroviral gene therapy trials aimed at treating children suffering from the X-linked severe combined immunodeficiency disease (X-SCID). Although the treatment was clearly beneficial, two out of the ten treated children developed a leukaemia-like disease as a result from integration of the retroviral genome close to an oncogene (Marshall, 2003). These adverse effects have raised much concern about the safety of integrating
10   vectors. As long as integrating gene therapy vectors insert the transgenes into random sites in the human genome they cannot be considered safe because of the risks associated with insertional mutagenesis.

The development of vectors capable of targeted transgene integration is under active investigation at the present. Hybrids vectors that are capable of targeted
15   transgene integration reach targeting efficiencies from 3 to approximately 40 percent: the baculovirus-adeno-associated virus vectors (Palombo et al., 1998) targeted transgene integration into the AAVS1 site in approximately 41% of the cases and the Ad-AAV hybrids with efficiencies ranging from 3% up to 35% (Recchia et al., 1999). An AAV vector designed to mediate site-specific integration by a transient Rep expression
20   reached targeting efficiencies of 42% (Satoh et al., 2000). There still is a need for more efficient and safe gene therapy vectors for targeted transgene integration.

3 SUMMARY

In vitro systems using viral preintegration complexes or purified IN (integrase protein) with short oligonucleotides have helped reveal important issues in the
25   integration reaction as well as possible ways to inhibit retroviral infection in target cells. The HIV-1 integrase has been fused to sequence-specific DNA-binding proteins to test the possibility of directing retroviral integration. Testing the activity of the fusion proteins *in vitro* has showed that the sequence specific proteins along with IN are capable of integrating substrate sequences at or close to the sites recognised by the
30   proteins fused to IN. A particular embodiment is a novel fusion protein consisting of

HIV-1 IN and a sequence-specific homing endonuclease I-*Ppo*I that could promote safe and targeted integration of gene therapy vectors. In addition, different DNA constructs were created for studying the integration mechanism of the novel protein *in vitro*. The homing endonuclease I-*Ppo*I recognises and cleaves its homing site present in the conserved 28S rDNA repeat of eukaryotes. While HIV-1 IN is designed to mediate the integration of retrovirus-like substrate sequences, I-*Ppo*I is designed to target the integration into the abundant rDNA locus in eukaryotic genomes. The aim was thus to create all the components required to test the integration activities and targeting abilities of a novel fusion protein consisting of HIV-1 IN and the HE I-*Ppo*I or I-PpoI's mutant form H98A. This aim included:

1- Creating the DNA constructs needed in expression of the IN-I-*Ppo*I and IN-H98A fusion

   proteins, as well as wt HIV-1 IN (control IN)

2- Designing and creating an LTR-flanked integration substrate

3- Generating an integration target plasmid containing the I-*Ppo*I recognition sequence

4- Production of the novel fusion proteins in bacterial hosts.

Some general aspects are: a polypeptide integrase, especially lentivirus integrase, and DNA-binding, especially with respect to human rDNA, activities; polynucleotides and vectors encoding it; its expression and production in a transformed host; compositions for administration comprising it; and its use in therapy, especially in targeted gene integration.

4 MATERIALS AND METHODS

4.1 General methods in DNA manipulation

4.1.1 PCR

PCR reactions were performed with a thermal cycler (PTC-200 Peltier Thermal Cycler, MJ Research) using the programs listed in Appendix II. Primers used in amplifying different templates are described in Appendix I. The genes for HIV-1 integrase, I-*Ppo*I and its mutant form I-*Ppo*Ih98a were amplified using *Pfu* –polymerase (MBI Fermentas). GW-PCR (subsection 4.3.1) was carried out using the FailSafe™

PCR PreMix selection Kit (Epicentre). In colony PCR (4.1.9) and other PCR reactions used to check different cloning constructs, DyNAzyme™ II DNA polymerase (Finnzymes) and its recommended buffer was used.

4.1.2 AGE and gel extraction

5 Throughout this study, gels for agarose gel electrophoresis (AGE) were prepared by dissolving 0,8 to 1 gram of agarose (Promega) in 100ml of TBE- or TAE-buffer (0,8 or 1% gels, w/v). Ethidium bromide (EtBr) was added to a final concentration of 0,4 μg/ml before casting. Gels were run on 70-120 volts (V) for 1 to 4 hours and the DNA was visualised with an UV- Transilluminator (M-20 UVP, Upland, CA, USA).

10 For DNA gel extraction, a 0,8 % TBE- or TAE gel was cast. DNA bands to be extracted were excised from the gel using a scalpel. The DNA trapped in the slices was extracted using a gel extraction kit (GenElute™ Agarose Spin Columns, SIGMA) according to manufacturer's instructions.

4.1.3 DNA purification, concentration and precipitation

15 DNA from different origins (from gel extraction, PCR, digestions etc) was purified and concentrated using the Wizard® Plus DNA Clean Up Kit (Promega) or alternatively Na-ac – alcohol- precipitation method (Sambrook & Russell, 2001) was used. DNA was usually eluted or dissolved in 30-50 μl of sterile endonuclease free water. The concentration of DNA and its purity in the eluate were determined 20 spectrophotometrically.

4.1.4 DNA concentration measurements

For DNA concentration measurements, samples were diluted in sterile water or TE (1:100). The concentration of DNA in the sample was spectrophotometrically determined by measuring the absorbance at 260 nm (and 280 nm), assuming that a 25 solution of 50 ng/ml gives an A_{260} value of 1. The purity of the sample was also determined from the ratio of the absorbance values at 260 and 280 nm (A_{260}/A_{280}) assuming that the ratio in a pure DNA solution is ~ 1,8 (Sambrook & Russell, 2001). Absorbance measurements were performed using an UV/Visible light spectrophotometre (Ultrospec 2000, Pharmacia Biotech).

4.1.5 DNA digestions

All digestions in this study were performed using restriction enzymes (REs) and their associated buffers from New England Biolabs (NEB) or MBI Fermentas. Digestions were carried out according to manufacturer's recommendations for the proper temperatures and buffer conditions for each enzyme. The DNA concentration in the digestion was adjusted to approximately 0,1 – 0,5 $\mu g/\mu l$. At least 1 U (unit) of enzyme was used for every μg of DNA, but less than 10% of the digestion volume consisted of the RE. Digestions were regularly carried out for 2 hours or over night (o/n).

When verifying plasmid preparations from different clonings by RE digestions (restriction analysis), the reaction was typically assembled as shown below, incubated for 1-2 hours at the appropriate temperature for each enzyme and run on an 1% TBE gel (w/v).

    2   μl plasmid DNA
    2   μl 10 X RE buffer
    2   μl 10 X BSA (for certain NEB enzymes only)
    0,5 μl of each RE
    x μl H$_2$O ad 20μl
    20μl

4.1.6 DNA ligations

Ligation of an insert into a linearised vector involves the formation of new phosphodiester bonds between adjacent 5'-phosphate and 3'-hydroxyl residues. In this study, ligation reactions were regularly performed using 5-15 U of the bacteriophage T4 DNA ligase (MBI), 1X T4 ligase buffer, variable amounts (50-150 ng) of digested vector and at least a 3 X molar excess of insert DNA over the vector DNA. All ligation components were mixed in a sterile eppendorf tube and sterile endonuclease free water was used to bring the reaction volume to 10 or 15 μl. Cohesive end ligations were usually incubated at room temperature (RT, 22°C) for 30-60 minutes and blunt end ligations at 16°C over night (o/n). Background ligation controls (water substituting for

insert DNA) were always carried out along the insert-ligations in order to determine the background level of non-recombinants.

4.1.7 Bacterial transformation

Dilutions of the ligation mix (approximately 500 and 250 pg DNA/µl) were used
5   to transform competent *E. coli* DH5α cells (Gibco) using 1 µl of each dilution per 30 µl of competent bacteria. Prior to transformation, T4 DNA-ligase was inactivated by incubating the ligation mixture at 65°C for 10 minutes. Transformation was carried out by the heat shock method as follows: the diluted ligation mix was mixed with DH5α cells in chilled eppendorf tubes, tapped gently and let stand on ice for 30 minutes. The
10   bacteria were subjected to a heat shock at +42°C water bath for 40 seconds and placed on ice for 2 minutes. 1x SOC (80µl) was then added on the bacteria and the cells were let recover in a shaking incubator for one hour (37°C, 250 rpm). Cells were spread on LB (Luria-Bertani) -agar plates supplied with the appropriate antibiotic for selection of transformed bacteria. When using blue/white colour selection of recombinant
15   pBluescript clones, 50µl of 2% X-Gal (5-bromo-4-chloro-3-indolyl-beta-D-galacoside) was spread on the surface of the plate and allowed to absorb for 30 to 60 minutes prior to plating the bacteria. 30 µl of IPTG (isopropyl thio-β-D-galactoside), an inducer for the *lacZ*-gene in pBluescript, was directly mixed with the bacteria upon plating.

In addition to the ligation control reaction, positive and negative transformation
20   controls were carried out simultaneously and given all the same treatments as for the ligation mix reactions. A positive transformation control comprised bacteria transformed with intact plasmid DNA and a negative control transformation with sterile water instead of any DNA. A positive control shows that the transformation reaction procedure works efficiently and a negative control is used to reveal possible DNA contaminations during
25   transformation. Agar plates were incubated at 37°C o/n.

4.1.8 Purification and analysis of plasmid DNA

The day following transformation, single white colonies were selected from LB plates and inoculated into 5 ml of 1 x LB supplemented with the appropriate antibiotic. The bacteria were grown in a shaking incubator (250 rpm at +37°C) for 12-16 hours to

increase the plasmid yield. On the following day, 2-5 ml of each bacterial culture was harvested and the plasmid-DNA was isolated using a mini-scale plasmid preparation kit (Wizard® *Plus* SV Minipreps DNA Purification System, Promega), according to the manufacturer's protocol. Plasmid preparations were identified by different restriction

5 enzyme digestions and agarose gel electrophoresis. Glycerol stocks were created from each succeeded clone by mixing equal volumes of bacterial o/n culture with 99+% glycerol (Sigma) and then stored at -20°C.

4.1.9 Colony PCR

Occasionally, cloning success was first verified by colony PCR instead of

10 verification of mini-scale plasmid preparations. Colony-PCR allows rapid detection of cloning success using bacterial colonies as templates for PCR. For each set of colonies to be tested, a PCR premix (master mix) was prepared (see below). Premix was aliquot in sterile eppendorf tubes and the bacteria serving as template were introduced by gently touching the colony with the tip of a pipette and then flushing it in the PCR-tube, being

15 careful not to pick the entire colony. Alternatively, a colony was suspended in 50μl of 1 X LB with the appropriate antibiotic, let grow for 2 hours in a shaking incubator (37°C, 250 rpm) and then 2μl of the culture was used for a modified colony- PCR reaction. With the latter procedure, numerous colonies could be screened simultaneously on a 96-well plate. PCR was carried out using the program DPESÄKE (appendix

20 II) and the reaction conditions listed below. The PCR-products were checked by AGE in a 1% TBE gel.

One 20 μl PCR-reaction was composed of:

15 μl (or 13 μl) PCR-grade water,

25 2 μl 10X Dynazyme PCR buffer (with $MgCl_2$)

1 μl DMSO (Dimethylsulfoxide, final concentration 5%),

0.5 μl dNTPs (10 mM),

0.5 μl primer T3

0.5 μl primer T7 (standard primers for pBluescript)

30 0.5 μl (1 U) DyNAzyme™ II DNA polymerase (Finnzymes)

(2 µl of bacterial culture or the picked colony as tempate)

20 µl

4.1.10 Creating DNA linkers by oligodeoxyribonucleotide hybridisation

Equal amounts of synthetic complementary oligodeoxyribonucleotides (1 nmol) were mixed in 100µl sterile endonuclease free water or TE. The solution was incubated at 95°C for 10 min to denature the strands and allowed slowly to cool to room temperature (RT) inside the block. Slow cooling ensured the strands to anneal correctly without any uncomplementary pairing or unwanted secondary structure conformations.

4.1.11 Sequence verification of recombinant constructs

Sequencing of all DNA constructs was performed with two automated DNA Sequencers (ALF and ALFexpress, Pharmacia) in the DNA Synthesis and Sequencing Facility of the A.I Virtanen Institute for Molecular Sciences, Kuopio. One microgram of template DNA was required for each sequencing reaction. Primers used for sequencing different samples are listed in table 4.6-I. All sequencing results were analysed using DNAMAN (version 5.2.9, Lynnon BioSoft).

4.2 Creating expression constructs of IN-I-PpoI, IN-H98A and cIN

The genes for HIV-1 integrase (Groarke, Hughes, Dutko; Hong et al., 1993; Sioud & Drilca, 1991), I-*Ppo*I (Muscarella & Vogt 1989, Muscarella et al., 1990) and I-*Ppo*Ih98a were first PCR-amplified (Arnheim & Erlich 1992) and then subcloned into the *Eco*RV digested pBluescript. Subcloned genes were fused to create the constructs pIP (IN-I-*Ppo*I) and pIH (IN-H98A; see figure 4.5-I for plasmid maps). These gene fusions were modified and transferred to an expression (Destination) vector pBVboostFG (Laitinen et al., unpublished results) using GATEWAY™ Cloning Technology (GibcoBRL®, Life Technologies).

4.2.1 Subcloning the genes for IN, cIN, I-Ppoi and H98A into pBluescript

The cDNAs for the HIV-1 IN, I-*Ppo*I and I-*Ppo*Ih98a genes were obtained from plasmids pLJS10, pCNPpo6 and pCNPpo6h98a, respectively. The plasmid pLJS10 was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: pLJS10 from Drs. JM Groarke, JV Hughes, and FJ Dutko. Plasmids pCNPpo6 and pCNPpo6h98a were a kind gift from Prof. Raymond J. Monnat Jr., University of Washington, Seattle.

Specific oligonucleotide primers for each gene (Appendix I) were designed on the basis of sequence information obtained from the plasmids' providers and the GenBank (accession #s: IN - AF029884; I-PpoI - M38131; I-PpoI(h98a) - 1CYQ_C) All primers except 3'IN (synthesized in TAG Copenhagen A/S) were synthesized using an ABI DNA Synthesizer in the DNA Synthesis and Sequencing Facility of the A.I Virtanen Institute for Molecular Sciences, Kuopio. New restriction enzyme sites for cloning purposes were included in the primers (appendix I).

PCR reactions were assembled as listed below and performed using the program DS2006 (Appendix II) for amplifying the genes IN, cIN, I-PpoI and H98A. Primers F992 and 3'IN were used to amplify IN; F992 and 3'cIN were used for cIN. Both I-PpoI and H98A were amplified using the same primer set (F987 and G7), since the point mutation of the latter did not coincide with the primer sequences. All the inserts were amplified using recombinant *Pfu* polymerase (MBI Fermentas) in a 40 µl reaction composed of:

27 µl PCR-grade water

4   µl 10X buffer for recombinant *Pfu* polymerase (+MgSO$_4$)

1   µl DMSO (Dimethylsulfoxide, c$_{final}$ 2,5%)

1   µl dNTPs (10 mM)

2   µl 5' primer (20µM)

2   µl 3' primer (20µM)

1   µl *Pfu* polymerase (2,5U)

2   µl template DNA (1:100 dilution of pLJS10, pCNPpo6 and pCNPpo6h98a)

40 µl

The PCR-reactions were checked by AGE running an aliquot of the reaction in a 1% TBE- gel. The remaining PCR product was then loaded on a 0.8% TAE- gel from which correct PCR-products were extracted. Bands corresponding to the expected PCR-products were gel extracted and purified as described under 4.1. The concentration of DNA and its purity in the eluate were determined spectrophotometrically (4.1.4).

PCR products amplified by *Pfu*-polymerase possess blunt ends, so blunt end cloning was used in subcloning the genes for IN, cIN, I-PpoI and H98A into the phagemid pBluescript (pBluescript® II SK-/+, Stratagene). pBluescript was linearised by digesting it with the blunt end cutter *Eco*RV (MBI). After 2 hours of incubation at +37° C, the digestion was checked by running an aliquot in a 1% TBE gel. To separate the linearised vector from the remaining amount of intact, undigested supercoiled DNA, the entire sample was gel extracted using GenElute™ Agarose Spin Columns and purified with the DNA Clean Up Kit.

Blunt-end ligation of IN, cIN, I-PpoI and H98A into the *Eco*RV digested pBluescript was performed by incubating each ligation mix (100ng of the vector, 2µl T4 ligase (10U), 1µl 10 X T4 ligase buffer, 100 to 125 ng of insert DNA (gel extracted and purified PCR products) and 3,5 µl H$_2$O) at +16°C o/n.

The ligation mixtures were transformed into competent *E.coli* DH5α cells as described in subsection 4.1.7. Bacteria were plated on LB plates supplemented with ampicillin (75µg/ml) and 50 µl X-gal. 30µl of IPTG was mixed with the bacteria prior to plating. Plates were incubated at 37°C o/n. The following day, white isolated colonies were selected from the LB$_{amp}$ plates, cultured in 5ml LB$_{amp}$ media and the culture was processed to plasmid preparations as described in section 4.1.8. Plasmid preparations were identified by restriction analysis and gel electrophoresis. The IN and cIN mini preps were digested with *Eco*RI (NEB). Correct preparations were further verified by digesting with *Pac*I and *Xba*I (NEB) (*Pac*I and *Sfi*I for cIN-clones), the sites of which were introduced into the IN-gene during PCR. Similarly, the plasmids containing I-PpoI or H98A were verified by a *Bam*HI-digestion, and a *Spe*I-*Sfi*I- double digestion (NEB).

All digestions were checked by AGE as described earlier. Correct clones in pBluescript were named pBIN for integrase, pBcIN for control integrase, pBIPpoI for I-PpoI and pBH98A for H98A.

4.2.2 Creating fusion genes from IN and I-PpoI/H98A

5   The fusion of the two genes (IN with either I-PpoI or H98A) was achieved by restriction enzyme based subcloning (Sambrook & Russel, 2001). First, the orientations of the inserts in pBluescript were deduced by different RE-digestions. The IN gene was found to lie in pBluescript in an orientation, where it could be detached from the plasmid by a *Xba*I-digestion from its 3' end, but stay attached to the plasmid from its 5' end. I-

10   PpoI and H98A were digested with *Spe*I and found to detach from their plasmids.

    pBIN was digested with *Xba*I to create cohesive ends (staggered cut ends, that can be ligated together) for the *Spe*I digested pBIPpoI or pBH98A fragments to be inserted. 2 µl of SAP (Shrimp alkaline phosphatase, 1un/µl) was added in the pBIN-digestion to minimise the background level, i.e. to prevent unligated pBIN plasmid from

15   self-ligating and transforming along the ligation mix. The I-PpoI and H98A inserts were digested from their plasmids using *Spe*I and the fragments were extracted from a 0.8% TBE gel. The purified I-PpoI and H98A fragments were individually ligated to the linearised, purified pBIN using the same conditions as described earlier (section 4.1.6). Ligation mixtures were transformed as described in section 4.1.7. Instead of extracting

20   the plasmid DNA from the resulting colonies, successful clones were verified by colony PCR (4.1.9). Colonies giving the expected result were inoculated in 5ml of LB_{amp} and processed to mini preps as in section 4.1.8. Mini preps were verified by *Pvu*II digestions and AGE. The orientation of the two cDNAs with respect to each other was deduced from the digested samples. Correctly orientated clones were named pIP for integrase-I-

25   PpoI fusion and pIH for integrase-H98A fusions. One clone of each construct was verified by sequencing (4.6).

4.2.3 Correction of mutations

    Sequencing of the clones pIP5 and pIH5 revealed a nonsense mutation in position 212 of the IN gene that was absent in the pBcIN5 clone. pIP5 and pIH5 were

30   repaired taking advantage of pBcIN5 as follows: a fragment of 527 bps was digested

from pIP5 and pIH5 with *Pac*I and *Afl*II and replaced by the corresponding but unmutated fragment from pBcIN5. The plasmid backbone from pIP5 and pIH5 was gel extracted and the insert fragment containing the mutation was discarded. The correcting fragment cut from pBcIN was isolated and purified and then used for ligation into the digested pIP and pIH –plasmids. Ligations, transformations and plasmid isolations were carried out as described earlier under 4.1. A successful clone could only be verified by sequencing. Clones named pkorj.IP3, pkorj.IH7 were sequenced with the primers listed in table 4.6-I.

12
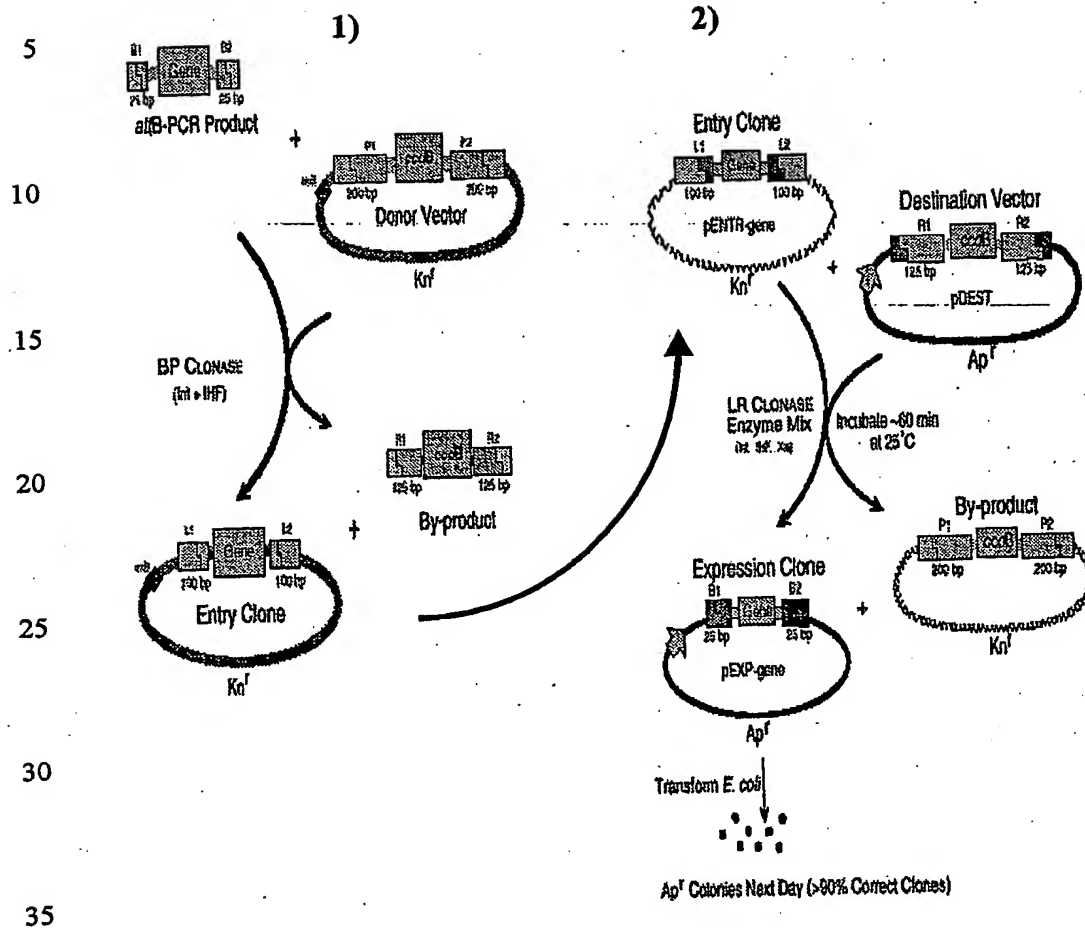
Figure 4.2.4-I: 1) Cloning a PCR product in Entry Vectors. The figure illustrates how cloning a PCR product using the BP Reaction is accomplished. The DNA sequence is first amplified with PCR, using primers containing the 25 bp *att*B sequences (Appendix I). This PCR product serves as the substrate for the BP Reaction. Recombination between the *att*B-PCR product of a gene and a Donor Plasmid (e.g. pDONR201) results in an Entry Clone of the PCR product, with kanamycin resistance. 2) The LR Reaction is a recombination reaction between an Entry Clone and a Destination (pDEST™) Vector, mediated by the GATEWAY LR CLONASE™ Enzyme Mix, a cocktail of recombination proteins. This reaction transfers DNA segments from the Entry Clone (pENTR™-gene) into a Destination Vector, to create an Expression Clone. The LR CLONASE Enzyme Mix that mediates the GATEWAY LR Reaction contains lambda recombination proteins Int (Integrase), Xis (Excisionase), and IHF (Integration Host Factor). In contrast, the BP CLONASE Enzyme Mix, which mediates the BP Reaction, comprises Int and IHF. (http://www.invitrogen.com/content.cfm?pfile='gwpcrcloning. htm')

4.2.4 Construction of expression vectors

To convert the fusion genes IN-I-PpoI and IN-H98A, as well as the c-IN clone, into constructs that could be expressed in bacterial, insect and mammalian cells, all the inserts were subcloned into new expression vectors. Recombination-based GATEWAY™ Cloning Technology (GibcoBRL®, Life Technologies) was used first to introduce the GW-PCR-modified genes into the Donor plasmid pDONR™ (Invitrogen) via a recombination based BP-reaction. From this plasmid the genes were again transferred via a LR reaction to a modified Destination Vector pBVboostFG (Laitinen et al., Unpublished results) that has the promoter elements necessary for transgene expression in different organisms. Polyhistidine tags (6 X His) were introduced to the 5' ends of the fusion genes. An illustration of the overall Gateway cloning technology is represented in figure 4.2.4- I.

4.2.4.1 Gateway (GW)-PCR

To create Entry Clones of pkorj.IP3, pkorj.IH7 and pBcIN5 by the GATEWAY-recombination technology, DNA elements (attB1 and attB2) were introduced in the 5' end of the genes by PCR. Primers (Appendix I, table I-2) were designed to contain the sequences attB1 and attB2 and the 5' primer was further modified to contain six extra histidine encoding codons to aid in future protein purification. PCR was performed with the FailSafe™ PCR PreMix selection Kit (Epicentre) using the kit included buffers A to L (buffered salt solutions containing dNTPs, various amounts of MgCl$_2$ and FailSafe PCR Enhancer with betaine), FailSafe PCR Enzyme Mix (DNA polymerases) and the program GW-2701 or GW-cIN (Appendix II). 1-3 ng of each template (pkorj.IP3, pkorj.IH7 or pBcIN) was used for a 50 µl PCR reaction. The products were gel-extracted from a 0.8% TBE gel as described before and concentrated by ethanol precipitation.

4.2.4.2 Construction of final expression vectors via BP and LR reactions

BP reactions were carried out for the plasmids pkorj.IP3, pkorj.IH7 and pBcIN5 according to the manufacturer's protocol and the products were transformed into E. coli DH5α cells (4.1.7). Plasmid preparations were made as described in 4.1.8. The resulting Donor-plasmid preparations (Entry clones) were verified by PvuII (MBI) digestions after which they served as templates for the subsequent reactions, the LR reactions. LR

reactions, where the transgenes become transferred from the Donor vectors into the Destination vectors (pBVboostFG in this study), were similarly carried out following the manufacturer's instructions and transformed into DH5α cells. Plasmid preparations from the resulting colonies were verified by *Pvu*II digestions. Gene constructs in the final Destination vectors were verified by sequencing using the primers G502, G550, G449 and G448 (Appendix I, Table I-3). Correct clones named pDIP2, pDIH1 and pDcIN1 were used for bacterial protein production (4.6).

4.3 Preparing the integration target plasmid (pPPOsite)

The sequence for the cleavage site of I-*Ppo*I endonuclease was adopted from previous studies using the enzyme (Mannino et al., 1999; Monnat et al., 1999; Ellison & Vogt, 1993; Argast et al., 1998). To prepare a plasmid that contained a single specific cleavage site for I-*Ppo*I, a double-stranded oligonucleotide containing the I-*Ppo*I recognition sequence was inserted into the *Eco*RV site of pBluescript. Two 5' phosphorylated 15-mer complementary oligonucleotide strands G515 and G517 (Appendix I, table I-4) composing the recognition site were annealed to form a double stranded site (I-Ppo-oligo). Equal amounts of the strands were mixed and hybridised as described in section 4.1.10.

pBluescript was digested with *Eco*RV as described earlier and treated with SAP (2 U), gel extracted and purified as described in 4.1. The linker I-Ppo-oligo was ligated to the linearised and purified plasmid using a 15 X molar excess of the insert (17.5ng) over the digested plasmid (225 ng) and the reaction mixture was incubated at +16 °C over the weekend (o/we). The diluted ligation mixture was transformed into DH5α cells with the heat shock method as described earlier. Mini scale plasmid preparations were processed from the cultures grown from selected bacterial colonies as in section 4.1.7 and verified by a digestion with I-*Ppo*I (Promega).

4.4 Creating a substrate for integration (pB2LTR+EGFP)

To generate an integration substrate for the *in vitro* study, a double stranded donor DNA resembling the viral long terminal repeat -ends was created. First, for both LTRs (3' and 5' LTR), two complementary and 5' phosphorylated 30-mer

oligodeoxyribonucleotides were synthesised on the basis of sequence information obtained from studies using *in vitro* experiments with HIV-1 integrase (Brin and Leis, 2002a). Equimolar amounts (100 pmol) of the 5' and 3' LTR's (G604+G605 and G569+G570, respectively; Appendix I, table I-4) complementary oligonucleotides were
5   annealed as before in 4.1.10.

pBluescript was digested with *Spe*I, extracted from a 0,8% TBE-gel and purified as described earlier. The linearised plasmid was then digested with *Kpn*I and treated with SAP (2 U), gel-extracted and purified. 200ng of the vector was used in a standard 15µl ligation reaction containing 2,5 µl of the 3' LTR preparation (25ng). Ligation was
10   carried out for 30 minutes at room temperature, after which the volume was increased to 20µl by adding 0,5 µl T4 ligase buffer (10X), 1µl T4 DNA ligase, 1µl H$_2$O and 2,5 µl of the 5' LTR preparation (25ng). Ligation was continued for another 30 minutes, after which the ligase was inactivated and the mixture transformed into DH5α cells (4.1.7). During the following days, light blue colonies were processed to plasmid preparations,
15   which were verified by a *Sca*I digestion. Before inserting the marker gene cassette between the LTRs, one of these pB2LTR-plasmid preparations (#1) was verified by sequencing.

The 2561 bp EGFP (enhanced green fluorescent) -cassette to be inserted between the LTR-sequences was *Sph*I digested from pBVboostFG+EGFP (Laitinen et al.,
20   unpublished results), gel-extracted and purified as described before. pB2LTR#1 (containing 5' and 3' LTRs) was also digested by *Sph*I and treated with SAP. The EGFP marker gene cassette was attached to the digested and purified pB1 in a 15 µl-ligation reaction containing 110 ng of the digested plasmid, three times its molar amount of the EGFP-insert (about 280 ng), 12,5 U T4 DNA ligase and 1,5 µl of 10X T4 DNA ligase
25   buffer. Ligation was first carried out for 1 hour at RT and continued in a refrigerator (+8°C) o/we. Diluted ligation mixture and its controls were transformed into DH5α-*E. coli* cells. Ten resulting white colonies were selected for plasmid preparations which were checked by a *Sca*I-digestion. Correct clones were named pB2LTR+EGFP (numbers 1-10) and verified by a *Sca*I digestion and sequencing.

4.5 Sequence verifications of recombinant constructs

The primers used in sequencing different plasmids are listed in Appendix I. Plasmids pIP5 and pIH5 (from 4.2.3), as well as their corrected versions pkorj.IP3 and pkorj.IH7 (from 4.2.4) were sequenced with the four primers listed in table 4.5-I.

5   pBcIN5, pPPOsite6 as well as the pB2LTR+E6 clone were sequenced with the primers T7 and T3. Expression clones pDIP2 and pDIH1 in pBVBoostFG were sequenced with the primers G502, G550, G448 and G449. pDcIN was sequenced with the primers G502 and G550. All sequencing results were analysed using DNAMAN (version 5.2.9, Lynnon BioSoft). A summary of all the DNA constructs except the expression plasmids

10   are illustrated in figure 4.5-I.

sequencing primers/ templates	G448	G449	T3	T7	G502	G550
pIP5	★	★	★	★		
pIH5	★	★	★	★		
pBcIN5	★	★	★	★		
pkorj.IP2	★	★	★	★		
pkorj.IH7	★	★	★	★		
pDIP2	★	★			★	★
pDIH1	★	★			★	★
pDcIN1					★	★
pPPOsite6			★	★		
pB2LTR1			★	★		
pB2LTR+E6			★	★		

Table 4.5-I: Primers used for sequencing different DNA constructs. Sequences of the primers are   listed in Appendix I.
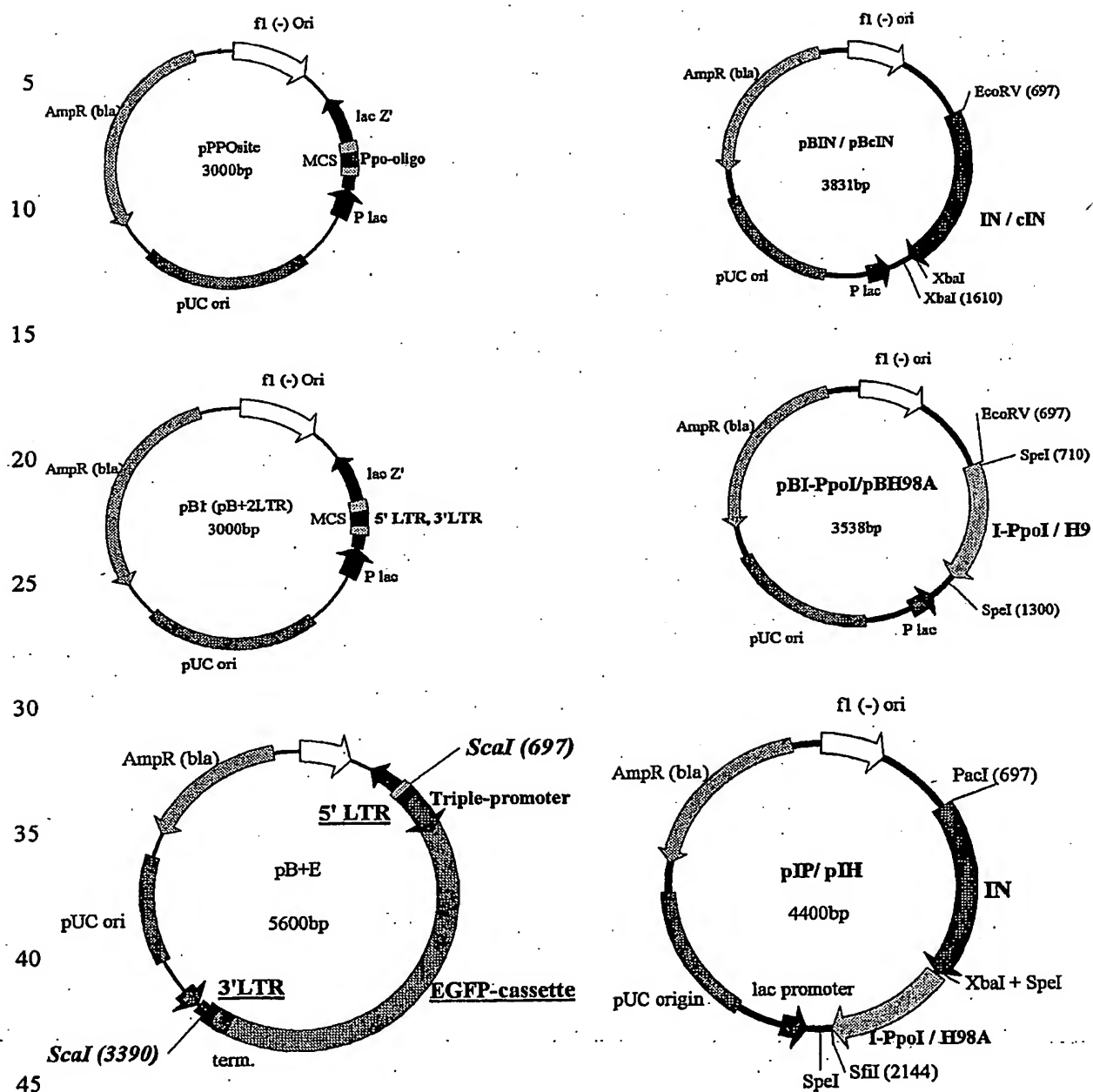
15

Figure 4.5-I: A summary of all the DNA constructs created before transferring the cDNAs of cIN, IN-H98A and IN-I-*Ppo*I (from the plasmids pBcIN, pIH and pIP, respectively) into the expression vector pBVboostFG.

4.6 Bacterial protein expression

Competent *E.coli* strains BL21 (DE3) (Stratagene) and BL21-AI™ One Shot® (Invitrogen life technologies, catalogue no. C6070-03) were used as expression hosts for
5   protein production. Time points of 0h, 1h and 3h post infection were taken during protein expression experiments and glycerol stocks were created from each tested clone. Protein production was analysed with SDS-PAGE (4.7.1) and Western Blotting (4.7.2).

4.6.1 Analytical scale bacterial protein production

Protein production was first carried out in the *E. coli* BL21 (DE3) strain. Bacteria
10   were transformed with the expression vectors pDIP2, pDIH1 and pDcIN1 and pINSD.His. The His-tagged HIV-1 IN encoding plasmid pINSD.His was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH from Dr. Robert Craigie: pINSD.His. Transformation was performed as follows: The bacteria from a glycerol stock or an untransformed colony were spread on an LB
15   plate the day before transformation in order to create a "fresh plate" and obtain viable colonies. The day following, isolated colonies were picked using a 10μl inoculation loop and suspended in 100μl of ice cold (+4 °C) CaCl$_2$. The suspended bacteria were incubated on ice for 15 minutes to render them competent for DNA uptake. DNA was added into the suspension (approximately 50-100ng, 0,5-1 μl), after which the mixture
20   was again incubated on ice for 30 minutes. Next, the cells were subjected to a heat shock at 42° C for 45 seconds and immediately placed on ice for two minutes. 450 μl of S.O.C was added on the cells after which they were let recover in a shaking incubator (37° C, 230 rpm) for 30-60 minutes. Various amounts of the transformation mixtures (50-150 μl) were plated on LBg plates (LB plates supplied with gentamicin) and incubated at
25   37°C o/n.

The following day, a few colonies were selected from each transformed construct and cultured in 5ml of 1 X LB supplied with gentamicin (7μg/ml) at 37°C, 250 rpm, until the OD$_{600}$ (optical density of culture at 600nm) reached 0,6 – 1,0. From these initial cultures, 200μl were used to inoculate 3,8 ml fresh LBg (1:20 dilution of the initial

culture). Untransformed BL21 (DE3) samples were processed as controls along the expression samples during all experiments.

The inoculated cultures were grown 1,5 to 3 hours at 37°C, 250 rpm, until they reached the mid-log phase (OD_{600} ≈0,4). Protein production was induced by adding IPTG to a final concentration of 1mM. Before induction, a 1ml sample (uninduced, 0h) was harvested from each culture. Additional 1 ml samples were collected 1h and 3 h (or every full hour until 4 h) post-induction. After harvesting, all samples were immediately pelleted by centrifugation (4000 x g, 2 min), suspended in 20-60µl sterile water and diluted in 4 X SDS-PAGE sample buffer. Samples were then heated to 95°C for 5 minutes to lyse the cells and stored at -20°C until analysed by SDS-PAGE. A part of the samples were not boiled at this point but stored immediately at -20°C.

Protein production was also tested in the *E. coli* strain BL21-AI™ One Shot® (Invitrogen). One or half a vial (50 or 25 µl) of BL21-AI™ was used for each transformation of the constructs pDIP2, pDIH1 and pDcIN1. Transformation was carried out according to the basic transformation procedure described in section 4.1.7. Three transformants (colonies) were selected from each construct and initial culture was grown as described for BL21 (DE3). From these cultures, 200µl were used to inoculate 3,8 ml fresh LBg (1:20 dilution of the initial culture). Untransformed BL21-AI samples were processed as controls. The inoculated BL21-AI- cultures were grown 1,5 to 3 hours at 37°C or at 30°C, 250 rpm, until the OD_{600} of the culture was approximately 0,4. Protein production was induced by adding L-arabinose (20% stock solution, Invitrogen) to a final concentration of 0,2% and IPTG to a final concentration of 1mM, 0,1mM or 0,01mM. Also, the impact of glucose in the growth medium was tested for production of the IN-I-PpoI protein (from pDIH). Glucose was added to a final concentration of 0,1% and all other steps were performed as previously. Expression cultures were usually grown at 37°C except when testing the impact of lower growth temperature (30°C) on protein degradation. Protein production samples were collected and processed as with BL21 (DE3) samples.

4.7 Characterisation of proteins

Samples from bacterial protein expression were prepared for SDS-PAGE immediately after harvesting. Before loading on the gel, samples were heated for 5 minutes at 95°C and placed on ice. The SDS-PAGE gels were blotted on nitrocellulose filters (Western blot) and the blotted proteins were identified by specific and sensitive antibody staining. Detection of the proteins with antibodies is based on alkaline phosphatase-activity conjugated-to-a-secondary-antibody (Blake et al., 1984).

4.7.1. SDS-PAGE

SDS-PAGE (Sodium dodecylsulfate-polyacrylamide gel electrophoresis) was performed using a Mini Protean II™ or a Mini Protean III™ device (BIO-RAD). First, a 10% running gel and a 4% stacking gel were cast (Appendix IV; Sambrook & Russell, 2001). 10μl of a molecular weight marker (SeeBlue ® Pre-Stained Standard, Invitrogen) and 20 to 50 μl of the samples and the controls were loaded in the wells of the stacking gel. The gel was first run with 100V for 10 to 20 minutes for the samples to migrate through the concentrating (stacking) gel. The volts were then raised to 180V (or 200V for Mini Protean III™ device) and the gel was run for additional ~60 minutes.

4.7.2 Immunoblotting (Western Blot)

The proteins resolved by SDS-PAGE were transferred to a nitrocellulose membrane (BIO-RAD Trans-Blot® Transfer Medium Pure Nitrocellulose Membrane, = 0.2 μm) using a BIO-RAD Mini Trans-Blot® device according to manufacturer's instructions. The transfer took place in cold Kodak buffer run with 100V for 1 hour, after which the blots were soaked in 0,5 M TBS buffer for 5 minutes. All proteins on the filter were visualised by staining the membranes with Ponceau S- dying solution (RT, 5 minutes, agitation) by covering the blots with the dye. The reaction was terminated by washing the membranes several times with distilled water. Membranes were then transferred to a blocking solution (5% non fat dried milk (NFDM) freshly made in TBS buffer) and agitated on a rotating shaker for 60 minutes at RT or o/n at 4°C.

Primary antibodies used in this study were (1) antisera to HIV-1 integrase peptide: aa 23-34 (NIH AIDS Research & Reference Reagent Program, catalogue #757)

and (2) monoclonal Anti-poly-Histidine-Alkaline Phosphatase Conjugate from mouse (SIGMA). The secondary antibody for (1) was Goat Anti-Rabbit IgG (H+L) −AP Conjugate (BIO-RAD). All the incubations at the antibody detection phase were carried out on a rotating shaker.

5   After blocking, blots were rinsed with TBS-Tween for 5 minutes and incubated in the primary antibody −solution (1:2000 both anti-IN and anti-Histag in 5% NFDM freshly made in TBS tween, 10 ml / membrane) for 1 hour at RT. When using anti-integrase staining, unbound primary antibody was washed with TBS-Tween (4 X 5 minutes). The blots were then incubated in the enzyme-linked secondary antibody

10   solution (goat anti-Rabbit IgG 1:3000 in 5% NFDM-TBS-Tween) for one hour at RT. Membranes were again rinsed with TBS-Tween (4 X 5 minutes) to eliminate unbound secondary antibody. When using anti-poly histidine antibody staining, blots were rinsed with TBS-Tween (4 X 5 minutes) after the first antibody probing and no secondary antibody was used.

15   Before colorimetric detection of proteins, membranes were equilibrated in APA-buffer by an incubation lasting approximately 5 minutes. Detection was achieved by incubating the filters in a NBT/BCIP substrate solution (Roche Diagnostics GmbH, Mannheim, Germany), 16μl/ml of NBT/BCIP in APA buffer, 5ml per filter at RT for 5 − 15 minutes. The colour reaction was terminated by washing the membranes several

20   times with deionised water. Filters were dried and the results of the staining were analysed.

5   RESULTS

5.1 Creating expression constructs of cIN, IN-I-PpoI and IN-H98A

25   The plasmid pLJS10 contains the gene for the HIV-1 integrase (IN) (Hong et al., 1993; Sioud & Drilca, 1991). Plasmid pCNPpo6 contains the gene for a eukaryotic intron encoded homing endonuclease I-*PpoI* from the myxomycete *Physarum Polycephalum* (Muscarella & Vogt, 1989; Muscarella et al., 1990). The I-*PpoI* gene (I-PpoI) in the plasmid pCNPpo6h98a has a mutation that replaces the amino acid residue

histidine in position 98 into alanine. Histidine98 (H98) in the enzyme's active site is critical for efficient catalysis of DNA cleavage and its mutation into alanine severely diminishes the endonuclease activity of the protein (Mannino et al., 1999). Therefore, throughout this work the H98A mutated and endonucleolytically inactive I-PpoI gene is

5 called H98A. Plasmids pLJS10, pCNPpo6 and pCNPpo6h98a were taken for use without preliminary verification of their sequences.

5.1.1 Subcloning the genes for IN, I-Ppol, H98A and cIN

The genes for HIV-1 integrase, I-PpoI and H98A were PCR amplified as described in section 4.2.1. The primers (Appendix I, table I-1) used for PCR

10 amplification were designed in a way to introduce new RE sites in the 5' and 3' ends of the genes. Two versions were created from the IN gene: one to be attached from its PCR-created 3'-XbaI-end to 5'-SpeI ended I-PpoI and H98A (IN) and another to become expressed without fusion partners (cIN; control IN). The PCR reactions were checked on a 1% TBE gel. The PCR products of the integrase gene were 870 bps long and the

15 products from the I-PpoI and H98A genes were 570 bps. (Figure 5.1.1-I)
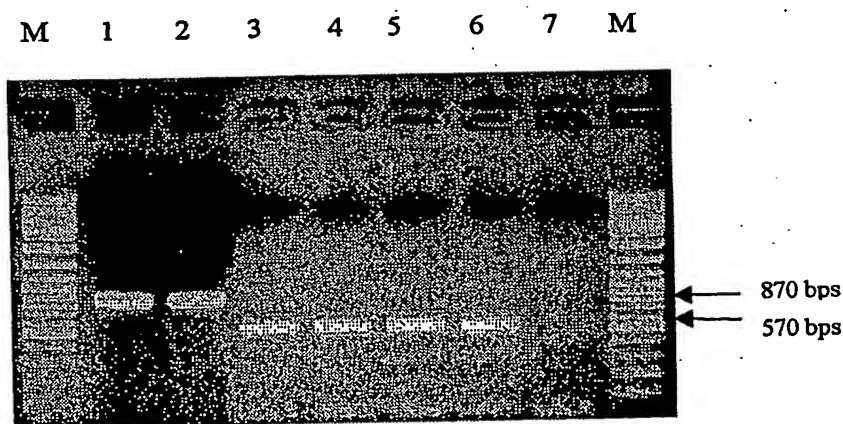
Figure 5.1.1-I: Analytical and preparative AGE of the PCR products. M: DNA size marker (GeneRuler™ DNA Ladder Mix, see Appendix III); lanes 1 and 2: IN PCR-products (870 bps); lanes 3-4: I-PpoI PCR-products (570 bps); lanes 5-6 H98A PCR-products (570 bps) lane 7: negative control of PCR.

In order to subclone the PCR products into pBluescript (Stratagene) the plasmid was digested with the blunt end cutter *Eco*RV, gel extracted and purified. Gel extraction of the digested vector usually strongly reduces the background level resulting from non-recombinant parental vector. SAP was not used because the blunt ended PCR products

5   created by *Pfu* polymerase lack 5' phosphates needed in ligating the inserts into the linearised plasmid. The linearised pBluescript for gel extraction was 2961bps long (Figure 5.1.1-II).
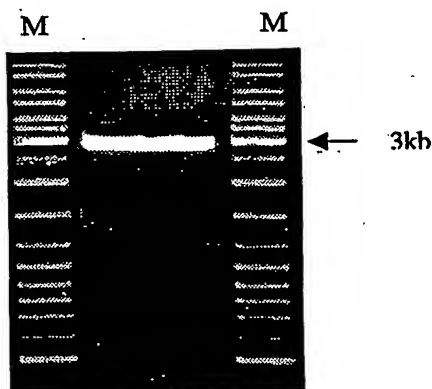
Figure 5.1.1-II: Preparative AGE of the *Eco*RV linearised pBluescript (2961 bps). M: DNA Ladder Mix

15   The PCR amplified inserts IN, cIN, I-PpoI and H98A were ligated to the *Eco*RV digested pBluescript as described in subsection 4.2.2. The choice of the ligase was the bacteriophage T4 DNA ligase (MBI) owing its ability to join both blunt-ended and cohesive DNA fragments (Sambrook and Russel, 2001). Ligation mixtures of IN, cIN, I-PpoI and H98A with the *Eco*RV digested pBluescript were transformed into *E. coli*

20   DH5α cells. Prior to transformation, ligation mixtures were heated to 65°C for 10 minutes to inactivate the ligase. The heat-inactivation step can increase the number of transformants by two orders of magnitude (Michelsen, 1995). Of the colonies developed, mini preps of all the constructs were verified by restriction analysis (4.1.5) and AGE as described earlier.

25   To screen the IN and cIN containing clones, mini preps were first digested with *Eco*RI. A successful clone was expected to appear as two bands of the sizes 465 and 3366 bps when visualised in a TBE-gel (Figure 5.1.1-III, lanes 3-5). Mini preps giving

the expected result were further verified by a double digestion using REs *Pac*I and *Xba*I or *Pac*I and *Sfi*I for cIN-clones. These sites were introduced to the insert ends in the PCR reaction. The correct clones could be visualised by the detachment of an 870 bp sized fragment corresponding to the IN and cIN genes. (Figure 5.1.1-III, lanes 7 and 8).
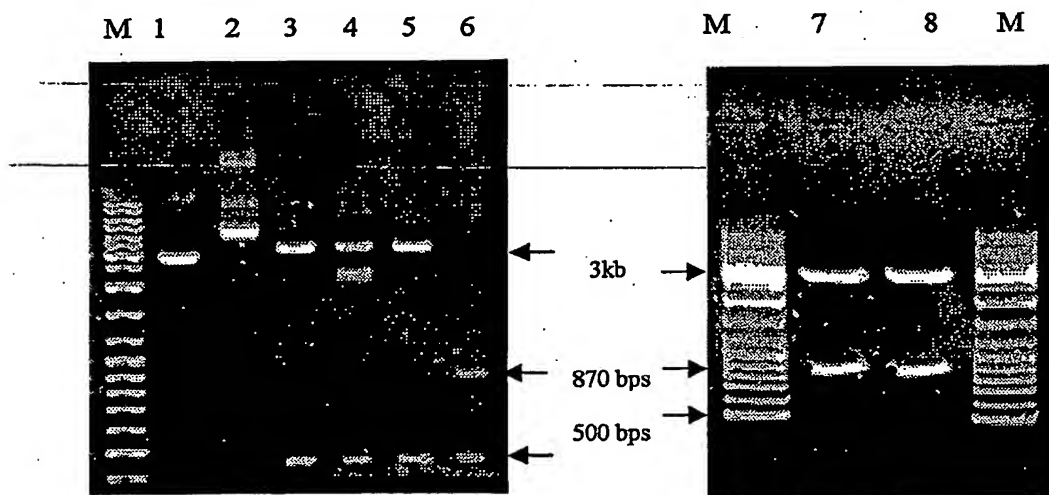
5

Figure 5.1.1-III: M: DNA molecular size marker (GeneRuler™ DNA Ladder Mix); lane 1: pBluescript/*Eco*RI; lane 2: undigested plasmid pBIN1; lanes 3-5: *Eco*RI digested pBIN mini preps; lane 6: *Eco*RI digested IN PCR product; lanes 7 and 8: *Pac*I and *Xba*I digested pBIN mini preps. Clones on lanes 3, 5, 7 and 8 are correct.

Similarly, the plasmids containing either I-PpoI or H98A were checked by digesting the plasmid preparations with *Bam*HI (Figure 5.1.1-IV). A correct clone was expected to appear as two bands of 336 and 3202 bps. Clones were further verified by a double digestion with the insert specific restriction enzymes *Spe*I and *Sfi*I, which produced the 570 bp sized I-PpoI or H98A insert. Correct clones of the inserts in pBluescript were named pBIN for integrase, pBcIN for control integrase, pBIPpoI for I-PpoI, and pBH98A for H98A. One of each correct clone was chosen to be used in creating the fusion gene constructs.
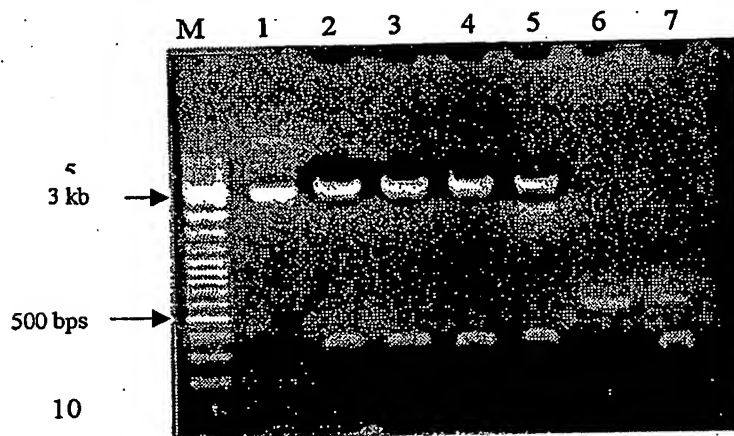
Figure 5.1.1-IV: Restriction analysis of pBIPpoI and pBH98A mini preps using *Bam*HI. M: DNA Ladder Mix; lane 1: undigested mini prep pBIPpoI 1,2; lanes 2-5: pBIPpoI mini preps digested with *Bam*HI; lane 6: undigested PCR product I-PpoI; lane 7: *Bam*HI digested PCR product I-PpoI. Clones on lanes 2-4 are correct.

5.1.2 Creating the fusion genes IN-I-*Ppo*I and IN-H98A

The IN gene residing in pBIN was fused to the I-PpoI and H98A fragments derived from pBIPpoI and pBH98A. First the orientations of the genes in their relevant plasmids were deduced by RE digestions. pBIN was digested with *Xba*I and its 5' end was found to stay attached to the plasmid while the gene's 3' end was freed along plasmid linearisation (Figure 5.1.2-Ib). This orientation was optimal, considering the fusion of the I-PpoI and H98A fragments into the integrase' 3' end. On the contrary, pBIPpoI and pBH98A clones were found to detach the inserted transgene when digested with *Spe*I, the enzyme producing compatible ends to the staggered cut created by *Xba*I (Figure 5.1.2-Ib). The detachment occurred because of another *Spe*I-site residing in the plasmid backbone, next to the inserts 3' end (figure 5.1.2-Ia). Fusions of I-PpoI and H98A with IN were thus achieved by ligating the *Spe*I cut "fusion partner gene" fragment into the *Xba*I linearised pBIN.

Ligation mixtures containing the *Xba*I linearised pBIN and the *Spe*I digested fragments of I-PpoI or H98A were transformed into DH5α cells as described in 4.1.7. Clones with the wanted gene fusion were first screened by colony PCR (4.1.9) instead of

directly proceeding to mini scale plasmid preparation from randomly selected colonies. A PCR product of ≈1610 bps (170 bp from pBluescript + 1440 from fused inserts) was expected from colonies having the IN gene fused to I-PpoI or H98A, whereas plasmids having only the IN gene inserted (the "background" clones) would give a product of

5   ≈1040 bps (170+870 bp; lanes 5,6,8 and 10 in Figure 5.1.2-II). Colonies giving the expected 1610 bp result (lanes 1-3, 7, 9 and 11 in Figure 5.1.2-II) were used to inoculate 5ml LB$_{amp}$ and the cultures were processed to mini preps as described before in 4.1.
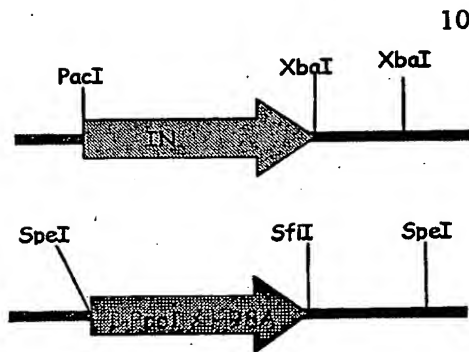
Figure 5.1.2-Ia: An illustration of the inserts' orientation in pBluescript. The SpeI and XbaI sites designed at the junction of the IN and I-PpoI / H98A genes are compatible and facilitate the joining of the fusion partner gene to IN.
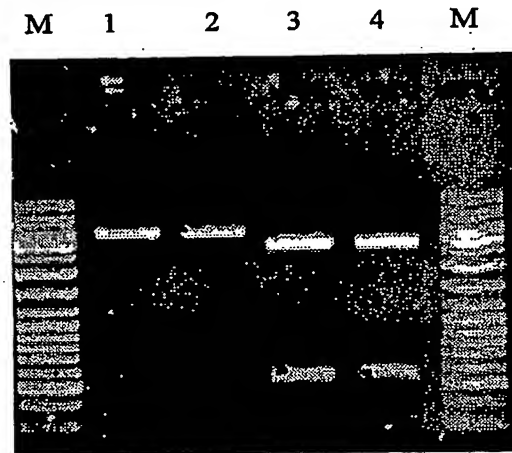
Figure 5.1.2-Ib: Testing the orientation of inserts in pBluescript. Lane 1: pBIN (1:7) /XbaI; lane 2: pBIN (1:3,4) /XbaI; lane 3: pBIPpoI (2:1,2) /SpeI; lane 4: pBH98A (3:7) /SpeI. M: DNA Ladder Mix
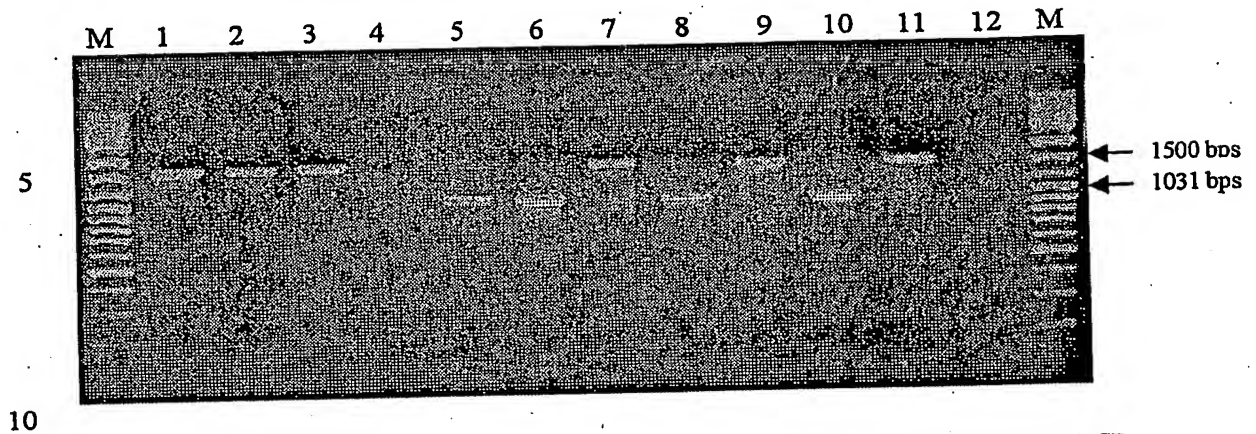
Figure 5.1.2-II: Colony PCR from IN-I-PpoI fusion clones. Each lane (1-11) represents the PCR product of a different bacterial colony. Lanes 1-3, 7, 9 and 11: bacterial clones with the wanted gene fusion; lanes 5, 6, 8 and 10: pBIN plasmid containing bacterial clones; lane 12: negative control of PCR; M: DNA Ladder Mix.

*Spe*I digested I-PpoI or H98A fragments could become ligated to the *Xba*I digested pBIN in two orientations (figure 5.1.2-IIIa and –IIIb) because both 5' and 3' ends of the fragments had compatible ends with the digested vector. Plasmid preparations containing the IN-I-PpoI or IN-H98A fusions were verified by *Pvu*II digestions. In addition to showing the presence of the gene fusion, *Pvu*II digestion also revealed the orientation of the inserts in respect to each other. Correct orientation of the genes fused to IN was essential considering the functionality of the novel proteins encoded by the fusion gene.
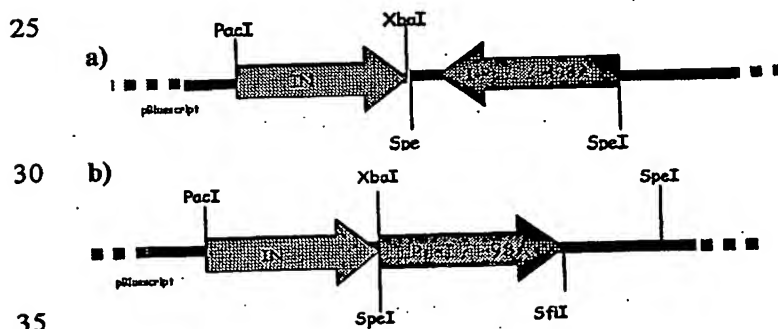

Figure 5.1.2-III: A) Correct orientation of inserts in respect to each other. B) Inversed orientation of IN and I-PpoI in respect to each other.

The desired fusion (thus I-PpoI's or H98A's 5' end ligated to the integrase' 3' end, Figure 5.1.2-IIIB) would result in four bands of the sizes 306, 728, 861, and about

5   2500 bps (Figure 5.1.2-IV) when digested with PvuII. The wrong orientation would yield bands of 306, 403 and 1186 and approximately 2500 bps. The correctly orientated clones were named pIP for integrase-I-PpoI fusion and pIH for integrase-H98A fusions. Clones number 5 were chosen from each construct (pBcIN included) to be verified by sequencing.
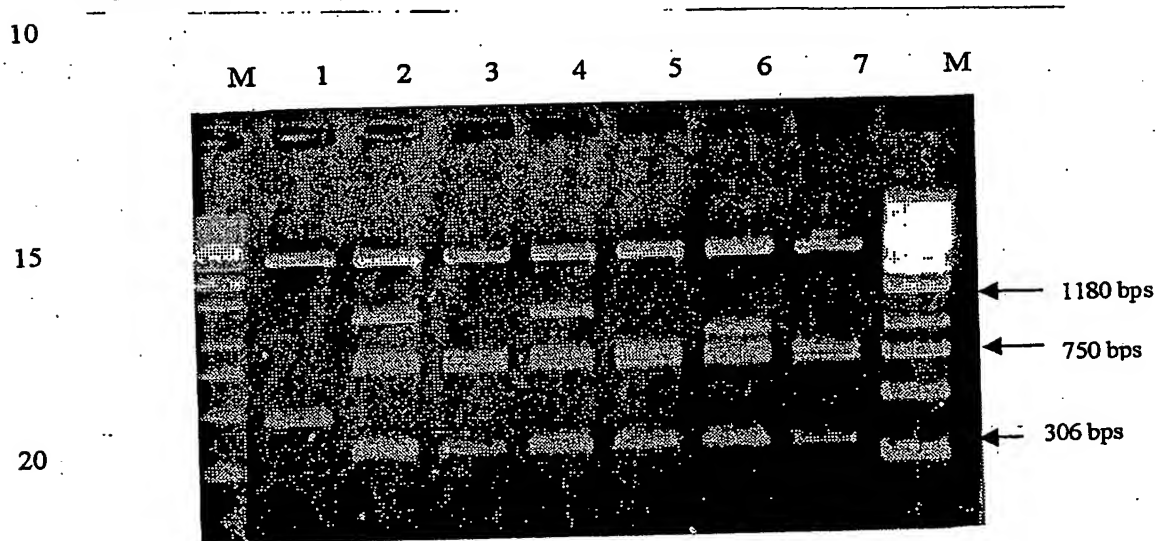
10

Figure 5.1.2-IV: Restriction analysis of pIP and pIH plasmid preparations by PvuII. M: DNA
molecular size marker (GeneRuler™ 1kb DNA Ladder, see Appendix III); lane 1: pBluescript/
PvuII; lane 2 to 7: pIH and pIP plasmid mini preps/ PvuII. Clones on lanes 3 (pIH5), 5 (pIP2)
and 7 (pIP5) are correct.

30 ### 5.1.2.1 Sequencing results

Plasmids pBcIN5, pIP5 and pIH5 were sequenced with the pBluescript's standard primers T7 and T3. Each sequencing reaction gave a two-way sequence that spun about 500 bp of the transgenes' sequence in pBluescript. Plasmids pIP5 and pIH5 were additionally sequenced with the primers G448 and G449 (table 4.6-I). These

35 primers were designed to anneal in opposite directions of the fusion gene's central region and they were used to amplify the sequence area not obtained using T3 and T7.

Sequencing results were analysed using DNAMAN (version 5.2.9, Lynnon BioSoft) by comparing the obtained sequences to the template sequences (AF029884 (HIV-1 IN), M38131 (I-PpoI) and sequences of the plasmids pCNPpo6 and pCNPpo6_h98a.

The sequence of the control integrase gene in pBcIN5 was found to be correct apart from a silent point mutation that did not affect the amino acid residue encoded by the mutated codon. Plasmids pIP5 and pIH5, on the contrary, were both affected by a nonsense mutation in position 212 of the IN gene. The mutation would cause the expressed protein to become severely truncated, which made correction of the mutated plasmids essential. Also the I-PpoI and H98A encoding parts in plasmids pIP5 and pIH5 differed from the sequence information obtained from the plasmid's (pCNPPo6 and pCNPpo6h98a) provider. To verify this result, second clones from the fusion constructs (pIP2 and pIH7) were sequenced with the same primers as pIP5 and pIH5 (table 4.5-I).

The parental plasmids pCNPPo6 and pCNPpo6h98a were sequenced with the primers G7 and F987 (Appendix I). The actual sequence of these plasmids was found to differ from the initial sequence information given by the plasmids' provider and from each other in more than the expected one position (His98-mutation in pCNPpo6h98a). The sequences obtained from sequencing pCNPPo6 and pCNPpo6h98a were more similar to the wild type sequence of I-*Ppo*I (M38131) than to the sequence information given by the plasmids' provider (table 5.1.2.2-I). Also, the sequence differences outside the expected H98A mutation were found not to have effects on the amino acids encoded by the differing codons. Thus, also the I-PpoI and H98A encoding sequences obtained from pIP5 and pIH5 were correct as they corresponded to the wild type protein's amino acid sequence.

CLUSTAL multiple sequence alignment

. pcnppo_h98a-cons = consensus sequence from sequenced parental plasmid (pCNPPo6_h98A)

M38131   = wt I-*Ppo*I sequence

pcnppo6_H98A   = sequence information obtained from plasmid pCNPPo6-h98a's
provider

pcnppo_h98a-cons CCCTTTTCTATTAGAACCGGATAACATCAACGGCAAAACGTGCACAGCAAGCGCGCTATG
M38131   CCCTTTTCTATTAGAACCGGATAACATCAACGGCAAAACCTGCACAGCATCGCCACCTATG
pcnppo6_H98A   CCCTTTTCTATTAGAACCGGATAACATCAACGGCAAAACCTGCACAGCATCGGCNCTATG
                  ** ********   *****
                                                           (1)   (2)
                                                           GCN-> His98A mutation

pcnppo_h98a-cons TCATAATACTCGATGCCACAATCCCTTGCACTTGTGCTGGGAGTCACTAGACGACAACAA
M38131   TCATAATACTCGATGCCACAATCCCTTGCACTTGTGCTGGGAGTCACTAGACGACAACAA
pcnppo6_H98A   TCATAATACTCGATGCCACAATCCCTTGCACTTGTGCTGGGAGTCACTAGACGACAACAA
                  ****

pcnppo_h98a-cons AGGCAGAAACTGGTGCCCGGGTCCCAACGGGGNATGTGTCCATGCGGTGGTTTGTTTAAG
M38131   AGGCAGAAACTGGTGCCCGGGTCCCAACGGGGGATGTGTCCATGCGGTGGTTTGTTTAAG
pcnppo6_H98A   AGGCAGTTTGACCAGCCCGGGTCCCAACGGGGGATGTGTCCATGCGGTGGTTTGTTTAAG
                  ******   ***************** ************************
                  (3)

Figure 5.1.2.2-I: Comparing the sequence of the plasmid pCNPPo6h98a to the wild type sequence information of the I-*Ppo*I gene and to the given sequence information by sequence alignment. Stars represent identical sequence, gaps in the consensus ("star"-) row indicate differences in sequence. The base difference in the first gap (1) has no effect on amino acid encoded by the affected codon. Differences in (2) are also silent apart from the His98 mutation presented as GCN in the sequence obtained from pCNPpo6h98a's provider (codon showed in bolded, wt encodes for His and mutated codon for Ala). Case (3) indicates a sequence difference of the parental plasmid compared to wt sequence of I-PpoI, but the actual sequence of the plasmid pCNPpo6h98a was found to be identical to the wt sequence. The plasmid pCNPpo6 only differed from the wt sequence in the expected His98 encoding triplet.

5.1.3 Correction of mutations in pIP5 and pIH5

The IN gene in pBcIN5 did not contain the nonsense mutation found in position 212 of the pIP5 and pIH5 fusion insert plasmids. This was partly due to the separate PCR-reactions used to amplify integrase genes for different purposes (having different

5   primers; Appendix I) and because of the possible heterogeneity of PCR products in one reaction.

The pIP5 and pIH5 plasmids harboring the nonsense mutation were repaired taking advantage of the unmutated IN gene in pBcIN. A fragment of 527 bps was detached

10   from the unmutated pBcIN5 with PacI and AflII. pIP5 and pIH5 were similarly digested with the two enzymes and treated with SAP. The gel extracted and purified pBcIN fragment (527 bp) was ligated to the digested, gel extracted and purified plasmid backbones of pIP5 and pIH5 (3880 bp) lacking the corresponding fragment. (Figure 5.1.3-I). A successful clone could only be verified by sequencing plasmid preparations.

15   To increase the probability of finding a successful clone, mini preps were only prepared from transformants having low levels of background (ligation) control colonies. First sequenced samples of corrected pIP5 and pIH5 (plasmids pkorj.IP3, pkorj.IH7) were found to lack the nonsense mutation and correction of the mutation had succeeded. The new fusion gene preparations were used in Gateway PCR.
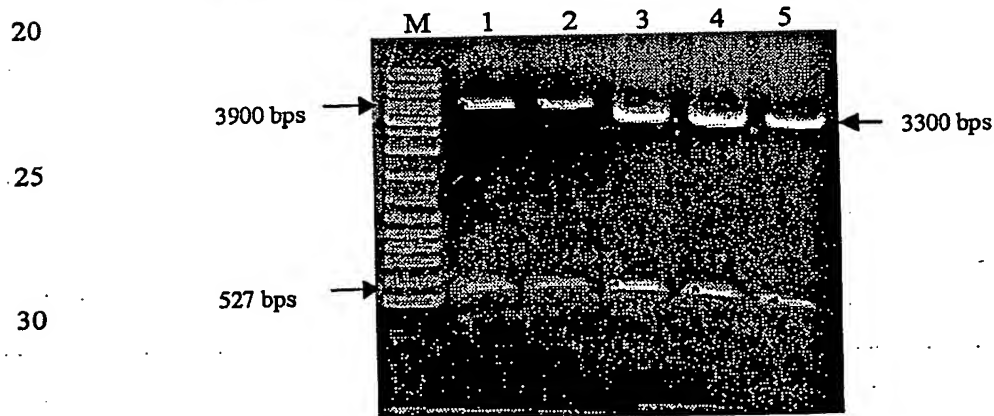
20

Figure 5.1.3-I: Correction of mutations in pIH5 and pIP5. M: DNA molecular size marker (GeneRuler™ DNA Ladder Mix); lane 1: pIP5/ *PacI*, *AflII*; lane 2: pIH5/ *PacI*, *AflII*; lanes 3-5: pBcIN5/ *PacI*, *AflII*. The marked bands were gel extracted and purified and used to create pIP and pIH plasmids without mutations.

5.1.4 Gateway (GW) PCR

Gateway PCR (4.2.5.1) was performed using the *att*B1/B2-sites and a 5' Histag
introducing primers (appendix I, table I-2) and the FailSafe™ PCR PreMix selection Kit
(Epicentre). The lengths of the primers G445 (5'), G238 (3' IN-Ppo and IN-H98A) and
G402 (3' cIN) were 86, 70 and 54 bps, respectively, so 110 or 104 bps were newly
introduced to the insert ends altogether. The expected size of the PCR products from
amplifying pkorj.IP3, pkorj.IH7 was thus 1440 + 110 = 1550 bps (Figure 5.1.4-Ia) and
for pBcIN5 974 bps (870 + 104) (Figure 5.1.4-Ib). PCR reactions were run on a 0,8%
TBE gel and the correct sized bands were gel extracted and purified.

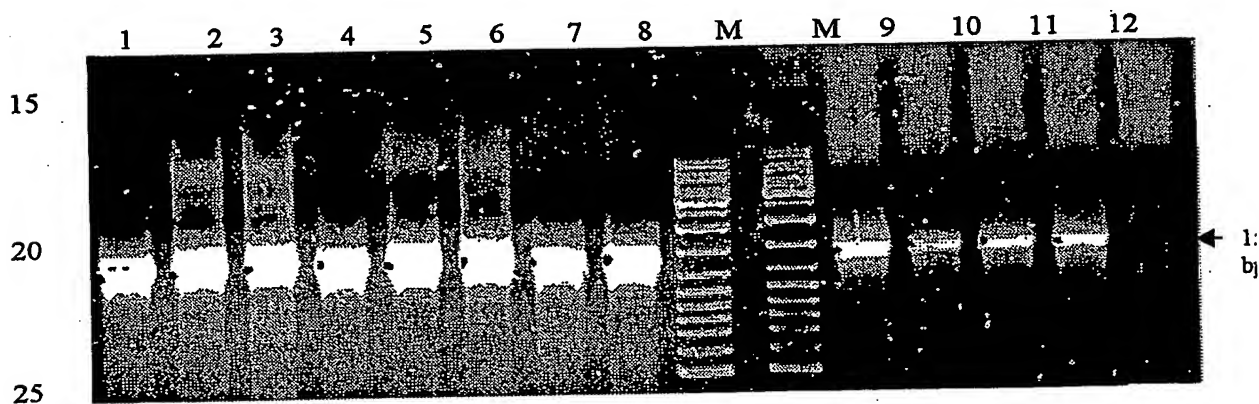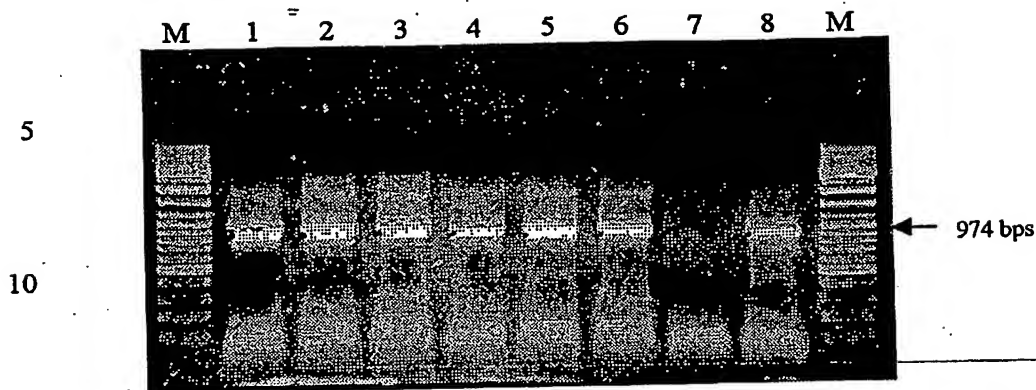

Figure 5.1.4-Ia: Preparative AGE of the GW-PCR products. Lanes 1-8: GW-PCR products
from amplifying pkorj.IP3 using FailSafe PCR PreMix™ buffers A-H and lanes 9-12 FailSafe
PCR PreMix™ buffers I-L. M: DNA ladder mix. The marked bands were gel extracted and
purified

| M | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | M |

← 974 bps

Figure 5.1.4-Ib: Preparative AGE of the GW-PCR products from amplifying pBcIN5. M: DNA Ladder Mix; lanes 1-8: GW-PCR products of pBcIN5 using FailSafe PCR PreMix™ buffers A-H. The marked bands were gel extracted and purified.

5.1.5 Construction of final expression vectors

BP reactions were carried out for the GW-PCR products of the plasmids pkorj.IP3, pkorj.IH7 and pBcIN5 following the manufacturer's protocol (GATEWAY™ Cloning Technology, GibcoBRL®, Life Technologies). The resulting Entry clones (pDONR™ plasmid preparations) were verified by PvuII (MBI) digestions. Correct Entry clones were identified by the presence of three bands of 2480, 810 and 560 bps in the case of plasmid preparations created from the IN-Ppo and IN-H98A fusion genes (pIPentry 1 and 2 and pIHentry 1-3 , Figure 5.1.5-I). The Entry clones bearing the cIN gene (pEntryCIN 1-3) were similarly restriction analysed using PvuII. A correct clone was visualised as the detachment of an ≈850 bp fragment (Figure 5.1.5-II), whereas an empty plasmid yielded a 602 bp fragment.
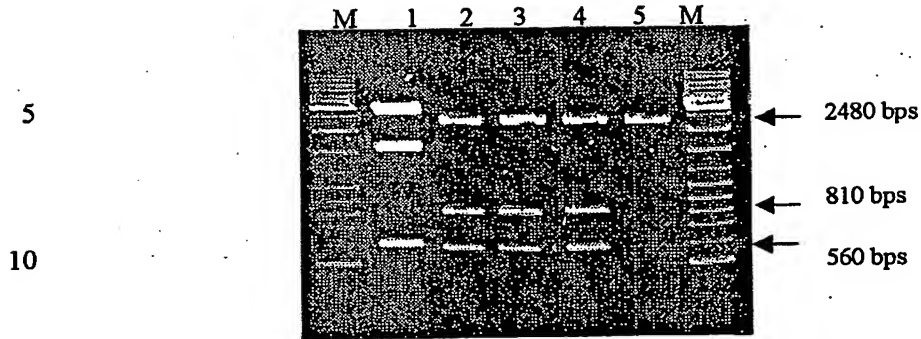
Figure 5.1.5-I: Restriction analysis of pIPentry (2 and 3) and pIHentry (1-3) plasmid preparations by *Pvu*II digestion. M: DNA Ladder Mix; lane 1: pIPentry 2/ *Pvu*II; lane 2: pIPentry 3/*Pvu*II; lane 3: pIHentry 1/ *Pvu*II; lane 4: pIHentry2/ *Pvu*II; lane 5: pIHentry3/ *Pvu*II. Clones on lanes 2-4 are correct.
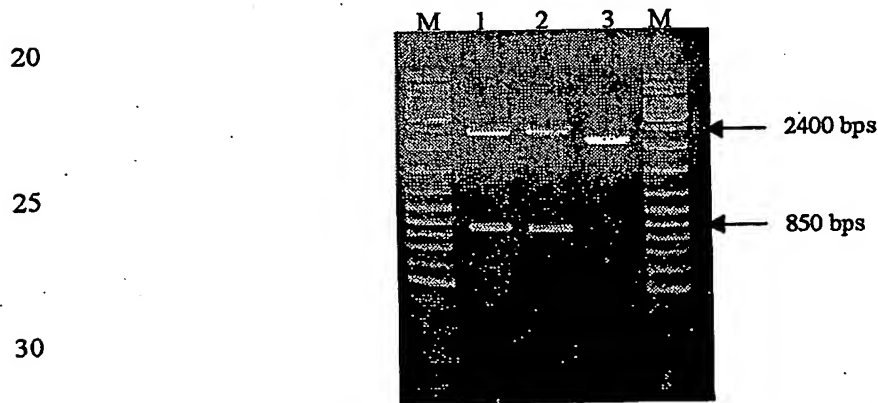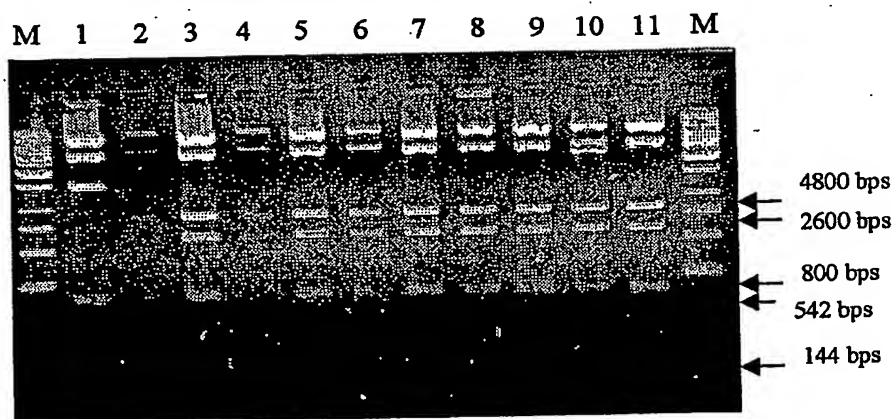

Figure 5.1.5-II: Restriction analysis of Entry clones bearing the cIN gene (pEntryCIN 1-3) by *Pvu*II digestion. M: DNA Ladder Mix; lane 1: pEntryCIN 1/ *Pvu*II; lane 2: pEntryCIN 2/ *Pvu*II; lane3: pEntryCIN3/ *Pvu*II. Clones on lanes 1 and 2 are correct.

Destination vectors that were created by the LR reaction between the Entry clone plasmids (pIPentry3, pIHentry1 and pEntryCIN1) and the Destination vector pBVboostFG (Olli Laitinen et al., unpublished results), were similarly carried out following the manufacturer's instructions. Mini preps were processed from the clones of the colonies resulting from the LR-reactions and verified by *Pvu*II digestions. The correct expression clones pDIP1-5 and pDIH1-5 were characterised by the appearance of 5 bands of the approximate sizes 4800, 2600, 800, 542 and 144 bps (Figure 5.1.5-III). The *Pvu*II digested intact plasmid pBVBoostFG was run next to the Destination plasmid digestions as a control (resulting in four bands of the sizes 4973, 2684, 1433 and 144

bps). The correct Expression plasmids containing cIN could be identified by the detachment of about 800 bps fragment upon digesting with *Pvu*II (Figure 5.1.5-IV). Clones giving the expected patterns were named pDIP, pDIH and pDcIN, the initial D standing for Destination plasmid.

5

Figure 5.1.5-III: Restriction analysis of expression vectors pDIP 1-5 and pDIH 1-5 by *Pvu*II
10   digestion. M: GeneRuler™ 1kb DNA ladder (see Appendix III); lane 1: pBVBoostFG/ *Pvu*II; lanes 2-6: plasmid mini preps pDIP 1-5/ *Pvu*II; lanes 7-11: plasmid mini preps pDIH 1-5/ *Pvu*II. All clones are correct except the clone on lane 10.

15
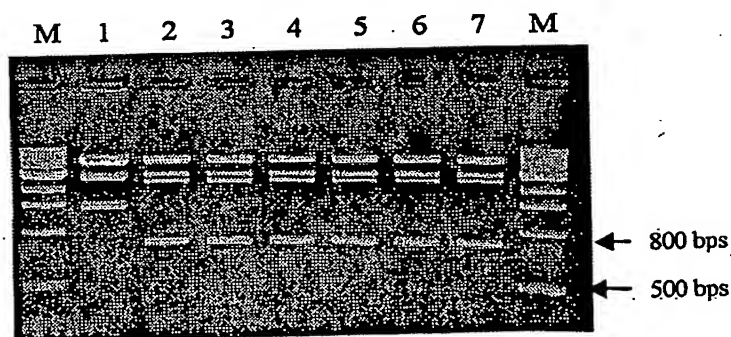
20   **Figure 5.1.5-IV**: Restriction analysis of expression vectors pDcIN 1-6 by *Pvu*II digestion. M: 1 kb DNA Ladder (NEB) (see Appendix III) Lane 1: pBVBoostFG / *Pvu*II; lane 2-7: pDcIN plasmid mini preps/ *Pvu*II.

Gene constructs in the final Expression vectors (clones pDIP2, pDIH1 and pDcIN1) were verified by sequencing with the primers listed in table 4.1.5-I. Sequencing results revealed the clones pDIP2, pDIH1 and pDcIN1 to be correct. These plasmid preparations were used for bacterial protein production (4.6)

5.2 Preparing the integration target plasmid (pPPOsite)

To prepare a plasmid that containes the cleavage site for I-PpoI, a double-stranded oligonucleotide containing the I-*PpoI* recognition sequence was inserted into the *EcoRV* site of pBluescript. The preparation of the oligonucleotide linker (I-Ppo-oligo) comprising the 15 bp recognition sequence, as well as other steps in creating the plasmid pPPOsite, are described in section 4.3.

The pPPOsite-plasmid mini preps were verified by restriction analysis using the enzyme I-*PpoI* (Promega). Clones bearing the I-Ppo-linker inserted became linearised upon digestion (Figure 5.2 -I). One clone (pPPOsite#6) was further verified by sequencing the plasmid with the primers T7 and T3. Sequencing revealed that the I-Ppo-linker was incorporated as two copies in the *EcoRV* digested pBluescript. Sequence of the linker was correct.
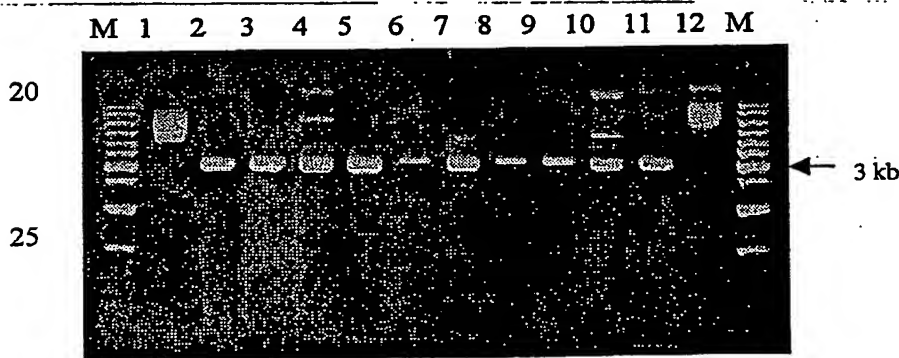

Figure 5.2 –I: Restriction analysis of pPPOsite plasmid mini preps by I-*PpoI* digestion. M: DNA Ladder Mix; Lane 1: undigested pPPOsite mini prep 1; lanes 2-11: pPPOsite mini preps 1-10 / I-*PpoI*; lane 12: a background plasmid (pBluescript)/ I-*PpoI*. All clones on lanes 2-11 are correct.

5.3 Creating a substrate for integration (pB2LTR+EGFP)

Requirements for the HIV-1 integrase substrate are, that it ought to be blunt ended, double stranded DNA, flanked with at least 15 base pairs of each of the viral 5' and 3' long terminal repeat ends (LTRs) (Brown, 1997). The terminal 20 base pairs of
5 the LTRs contain the crucial sequences needed for integration by HIV-1 IN, most importantly the dinucleotides 3'CA two bases away from each LTR's end (Brin & Leis, 2002a)

    The LTR ends were designed to be easily clonable to pBluescript and later to allow the insertion of a marker gene cassette between the LTRs (see figure 5.3-I). The 5'
10 LTR was designed to fit to KpnI-digested pBluescript from its 5' end and to the other LTR from its 3' end. The 3' LTR was designed to be ligated to pBluescript from its 3' end's SpeI site and to the 5' LTR from its SphI site. In addition, extra endonuclease cleaving sites were introduced in both the LTR-sequences: ScaI sites partly embedded in LTR sequences would allow release of the LTR flanked blunt ended EGFP cassette
15 bearing the crucial CA dinucleotides in both its 3' ends, two nucleotides away from the LTR termini.

20

```
                              C AGT ! ACTGGA AGG GCT AAT TCA CT GCATG
           KpnI-             CATGG TCA !TGACCT TCC CGA TTA AGT GA C          - SphI
25                                ScaI

                              C TGT GGA AAA TCT CTA GCAGT !ACT A
           SphI    -          GTACG ACA CCT TTT AGA GAT CGTCA ! TGA TGATC    - SpeI
30                                              ScaI
```
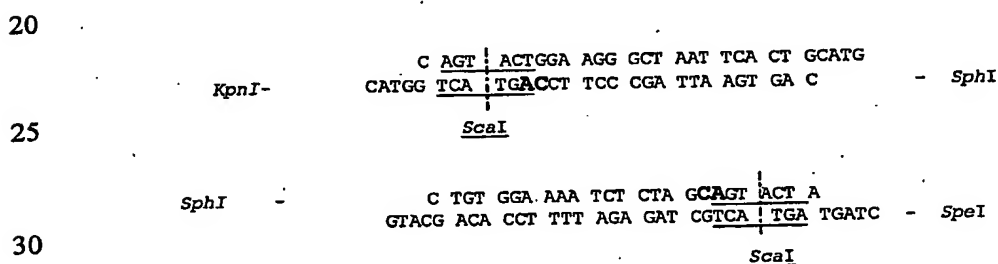
Figure 5.3.-1: Structure of the LTRs after hybridisation. The upper sequence presents the 5' LTR and the lower corresponds to the 3' LTR. The restriction site of ScaI is underlined and the
35 CA-dinucleotides in each 3' end are bolded.

    The preparation of the 5' and 3' LTRs, as well as their ligation into the KpnI and SpeI digested pBluescript is described in section 4.4. Mini preps were processed from light blue colonies that most likely had the short LTR sequences (that only partially
40 hindered the read-through of the α-fragment gene) inserted in the MCS. Plasmid

preparations (pB2LTR #1-3) were verified by restriction analysis using *Sca*I (MBI). Insertion of the two LTRs created two restriction sites for *Sca*I in addition to the site already residing in pBluescript. A correct clone having both the LTRs was thus identified by the detachment of an 1160 bps long fragment from pBluescript (the

5   sequence between the LTRs could not be detected due to its shortness) (Figure 5.3-I). Before inserting the marker gene cassette between the LTRs, a pB2LTR-plasmid preparation (pB2LTR#1) was verified by sequencing. Sequencing results obtained with the primers T7 and T3 showed the plasmid to be as expected, i.e. having both the 5' and 3' LTRs correctly inserted.
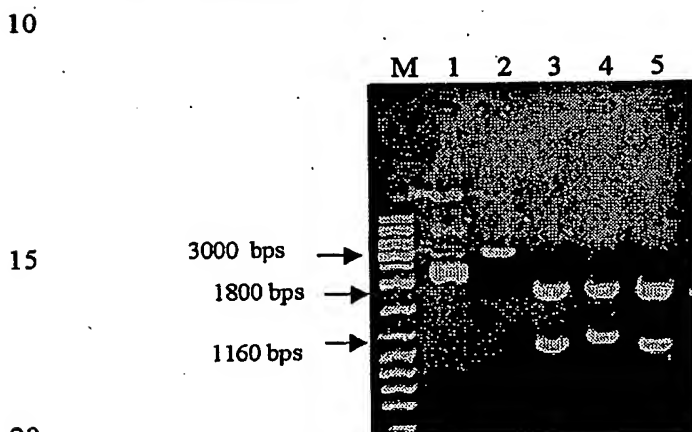
10

Figure 5.3-I: Restriction analysis of pB2LTR plasmid preparations (#1-3) by *Sca*I. M: DNA Ladder Mix. Lane 1: undigested pB2LTR#1; lane 2: pBluescript / *Sca*I; lane 3: pB2LTR#1 / *Sca*I; lane 4: pB2LTR#2/ *Sca*I; lane 5: pB2LTR#3/ *Sca*I. Clones on lanes 3 and 5 are correct.

25   The triple promoter containing EGFP marker gene cassette from pBVBoostFG-EGFP (Laitinen et al., unpublished results) was cut using *Sph*I (Figure 5.3.-II). The cassette was introduced into the pBluescript construct bearing the 5' and the 3' LTRs (pB2LTR#1) as described in 4.4. Insertion of the 2620 bp cassette could be verified by digesting mini preps with *Sca*I, which was designed to free the blunt ended EGFP

30   cassette. A correct clone was identified by the detachment of the EGFP-marker gene cassette (about 2620 bp) and two bands of about 1100 and 1800 bps resulting from the plasmid backbone. Restriction analysis showed 8/10 tested mini preps to be correct (Figure 5.3-III). Sequence of the EGFP cassette was assumed to be correct because it had not been modified after its sequence verification in pBVBoostFG. Orientation of the

cassette in respect to the LTRs was not crucial because the cassette itself carried the promoter elements needed for the expression of EGFP.
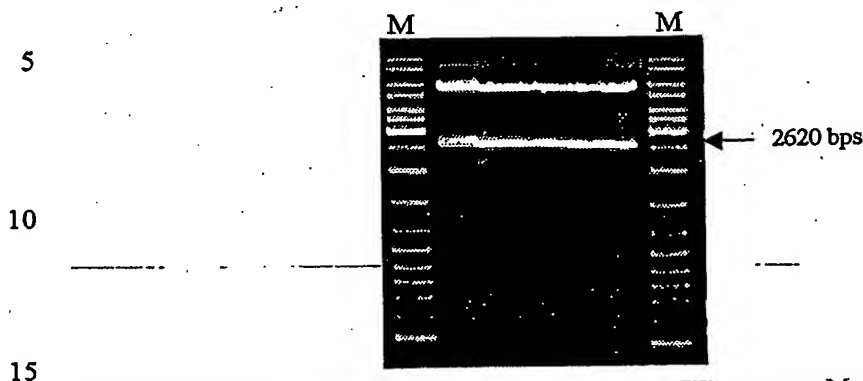

Figure 5.3-II: Preparative AGE of the *Sph*I digested EGFP cassette. M: DNA Ladder Mix.
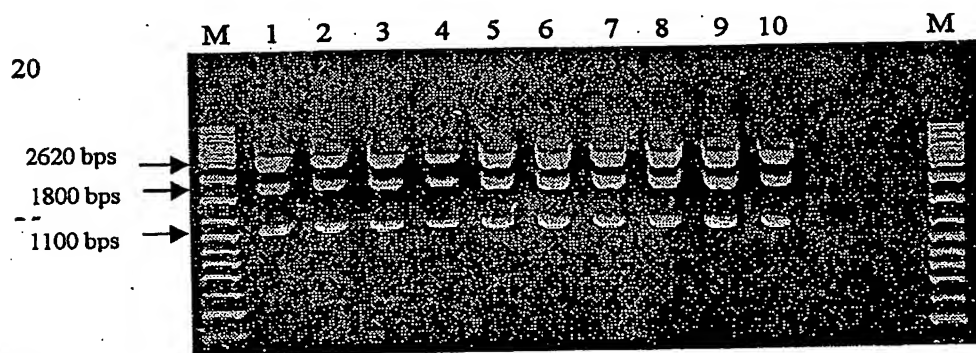
Figure 5.3-III: Restriction analysis of pB2LTR+E plasmid preparations. M: DNA Ladder Mix. Lanes 1-10: mini preps p2LTR+E #1-10 / *Sca*I. Clones on lanes 2 and 3 are not fully correct (an extra band of about 500 bps), all others are correct.

5.4 Bacterial protein production and analysis

The fusion genes IN-I-PpoI and IN-H98A were created in order to study the functionality of the novel fusion proteins they encode for, and to compare it with the activity of the native HIV-1 integrase (clone HXB2, cIN in this study). The genes were transferred into Destination vectors (pBVBoostFG) to be able to produce the proteins. Protein production experiments were carried out in bacterial cells using the *E.coli* strains BL21 (DE3) (Stratagene) and BL21-AI™ One Shot® (Invitrogen) as expression hosts.

5.4.1 Sample harvesting and preparation for SDS-PAGE

The bacteria were transformed with the expression vectors pDIP2, pDIH1 and pDcIN1 as described in 4.1.7. As a positive control for expression studies, also the plasmid pINSD.His (Engelman and Craigie 1992, Craigie et al 1995) was transformed into the expression host bacteria. Bacterial colonies were grown in a shaking incubator until the cell density reached the OD_{600} value of 0,6 – 1. At this stage, a part of the culture was inoculated in fresh LB_g media to compose the actual expression culture. Glycerol stocks were created from the remaining initial cultures of each clone. The expression culture was grown until the cell density reached the mid-log phase (OD_{600} ≈0,4) in order to maximise the capacity of the bacteria to overexpress the wanted proteins. As a control for Western blot, untransformed bacterial samples were processed along the actual expression (and positive control) samples.

Protein production in *E.coli* BL21 (DE3) cell culture was induced by adding IPTG (100mM) to a final concentration (c_{final}) of 1mM. Addition of IPTG induces the expression of the RNA polymerase T7 (T7RNAP) from the lacUV5 promoter. Expression of the T7 promoter can however be leaky, i.e. be "on" also in absence of the inducer in this cell. This can lead to uninduced expression of heterologous proteins, which may represent a disadvantage if the protein product is toxic. In effect, the gene product of pDIP2 was observed to hinder the growth of transformed bacteria on LB plates, as well as growth of bacterial cultures in LB media. Strict expression control was therefore a relevant issue in bacterial protein production.

BL21-AI™ One Shot® cells were induced with the addition of IPTG (c_{final} 1-0,01 mM) and L-arabinose (c_{final} 0,2 %). L-arabinose induction was needed to induce expression of the T7RNAP from the *ara*BAD promoter (P_{BAD}) in BL21-AI (Lee 1980, Lee et al., 1987). Because T7 RNAP levels can be tightly regulated in BL21-AI, the strain was thought to be especially suitable to express possibly toxic genes. Various amounts of IPTG were screened to study the effects of different induction levels on the stability of the overexpressed proteins. Also different growing conditions (lower temperature, addition of glucose) were tested for the same purposes.

Both the strains BL21 (DE3) and BL21-AI are deficient of the proteases *lon* and OmpT (outer membrane protease) which cause degradation of expressed heterologous

proteins in bacterial cells (Grodberg & Dunn, 1988; Studier & Moffat, 1986). Lack of these proteases thus reduces possibility of protein degradation in heterologous protein expression studies.

5   Time points of 1h, 2h, 3h and 4h (or some of them) were taken during expression experiments with both strains to define an ideal time for heterologous protein expression and stability. Samples were also taken from the cultures before induction (0h, UI) to assess the level of basal protein expression. All samples were harvested and prepared as described in section 4.7. Sample boiling immediately after cell pelleting was important in order to lyse the cells and inactivate the host's intracellular proteases. Part of the

10   samples obtained from BL21 (DE3) transformants were frozen right after sample preparation (cell pelleting and addition of SDS-PAGE sample buffer) and boiled only before loading on the SDS-PAGE gel. Protein production was analysed by SDS-PAGE (4.7.1) and Western Blotting (4.7.2).

5.4.2 SDS-PAGE and Western Blotting

15   Denaturing SDS-PAGE was performed as described in 4.7.1. The proteins were blotted onto nitrocellulose membranes during Western blot and stained as described in 4.7.2. Staining with the poly-histidine recognising antibody revealed the sizes of fragments bearing the N-terminal 6XHis fusion (Figure 5.4.2-II, right half of the blot) whereas staining with the HIV-1 integrase specific antibody revealed all fragments

20   bearing the N-terminal region (amino acids 23-34) of the IN protein. An I-*PpoI* specific antibody was not available for use, but the size difference of the proteins derived from pDIP2 and pDIH1 expression cultures compared to pDcIN derived proteins revealed the presence of correct fusion proteins (e.g. Figure 5.4.2-II). The protein product expressed from pDcIN was expected to be 32 kDa (as the monomeric wt HIV-1 integrase) and

25   those of the IN-H98A (pDIH) and IN-I-PpoI (pDIP) fusions 50 kDa (32 kDa + I-PpoI or H98A 18 kDa per monomer). A positive control for the immunoblot staining was prepared from the plasmid pINSD.His in the same way that expression samples from pDcIN1, pDIP2 and pDIH1 were processed. This His-tagged control reacted with both antibodies used in this study and revealed the size of the cloned integrase and the

30   presence of the poly-histidine tag in the protein's N-terminus.

Immunoblots derived from the expression samples in BL21 (DE3) showed the presence of expected protein bands resulting from specific antibody interactions in the staining procedure (Figure 5.5.2-I). In addition to the correct sized gene products of pDIH (Figure 5.4.2-I, lanes 2-4; correct size is marked with a black arrow), pDIP (Figure 5.5.2-I, lanes 5-7), and pDcIN (not shown for BL21(DE3), but see figure 5.4.2-IV for BL21-AI), numerous smaller bands were present in the blots. Proteins were thought to be subjected to degradation at some point of the expression pathway, possibly due to the strong basal expression from the T7 promoter and the consequential large quantity of the heterologous proteins in the cells. Expression of the heterologous proteins occurred already at the uninduced state of the BL21 (DE3) culture (Figure 5.4.2-I, lanes 2 and 5), which was thought to have contributed to protein instability.

Marked inhibition of bacterial growth was observed in cultured *E. coli* BL21 (DE3) cells when transformed with the expression plasmid pDIP2 (Figure 5.4.2-I, lanes 5-7). First, it was difficult to obtain any colonies when the bacteria were transformed with pDIP2. Secondly, both the initial culture as well as the expression culture of the pDIP2 transformed BL21 (DE3) cells grew very slowly if at all. Finally, the cells that could be extracted from the induced pDIP2 transformed BL21 (DE3) cultures had very little amounts of the expressed IN-I-*Ppo*I proteins (Figure 5.4.2-I, lanes 5-7).

To test whether the expression products of pDIP2, pDIH1 and pDcIN1 would be more stable when expressed in a strain more suitable for expression of toxic proteins, the experiments were repeated using the *E.coli* strain BL21-AI (Figures 5.4.2-II to 5.4.2-V). Immunoblotting results showed lower levels of basal expression in BL21-AI than in BL21(DE3) (Figure 5.4.2-III, lanes 2 and 3; Figure 5.4.2-IV, lanes 2-4; Figure 5.4.2-V, lanes 1-3 – compare e.g. with lane 2 in figure 5.4.2-I), but no clear enhancement in protein stability could be observed (Figures 5.4.2-II to 5.4.2-V). Different levels of IPTG induction did not change the expression levels from DcIN1 and pDIH1, nor did they affect protein stability (Figures 5.4.2-IV and 5.4.2-V).

Expression of the IN-I-PpoI protein from pDIP2 was again hindered as already observed with pDIP2 infected BL21 (DE3) cells. The protein product was thus thought to be toxic for the expression hosts used in this study. Addition of glucose (0,1 %) to the LB plates and the growth media was tested with pDIP2, because it may prevent

problems associated with the basal level expression from a toxic gene in BL21-AI cells [Manual for BL21-AI cells (Invitrogen) and references therein]. Glucose addition was found to have strong suppressive effects on the expression levels of the protein, but no significant improvement was observed on protein degradation (Figure 5.4.2-III, lanes 4-
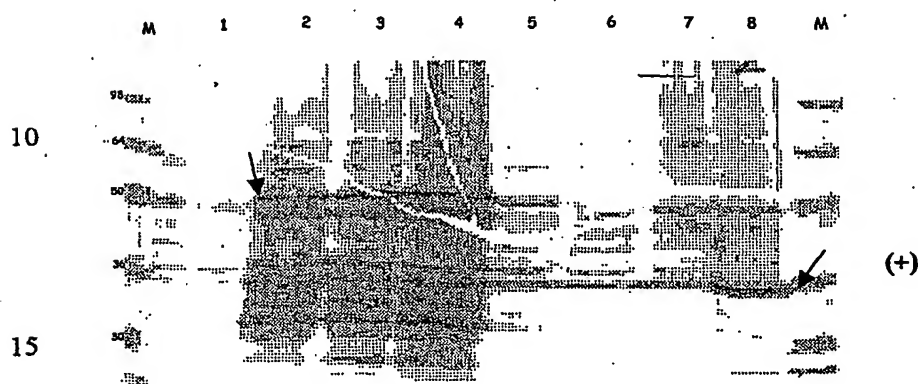
5 7).

Figure 5.4.2-I: Immunoblot from protein expression in pDIH1 and pDIP2 transformed BL21
20 (DE3) cells. The blot was stained with anti-integrase specific antibody. M: Protein molecular weight standard (SeeBlue® Pre-Stained Standard, Invitrogen) Lane 1: BL21- control (untransformed cell extract); Lane 2: pDIH1 uninduced (UI); Lane 3: pDIH1 1h; Lane 4: pDIH1 3h; Lane 5: pDIP2 UI; Lane 6: pDIP2 1h; Lane 7: pDIP2 3h; Lane 8: positive control (+) pINSD.His 4h. The specificity of the antibody (anti-HIV-1 integrase) can be noticed from the
25 unstained cell control (lane 1) and from the staining of the 32 kDa sized integrase protein in lane 8 (+; migrating near the 36 kDa size marker). Strong basal expression levels can be seen by comparing the expression intensities of lane 2 (pDIH1 UI) to lane 3 or 4 (pDIH1; 1 and 4 hours after IPTG induction, respectively). Low expression levels of pDIP2 transformed cells indicate toxicity of the protein product (lanes 5-7). Black arrows show the expected sizes of the
30 expressed proteins.
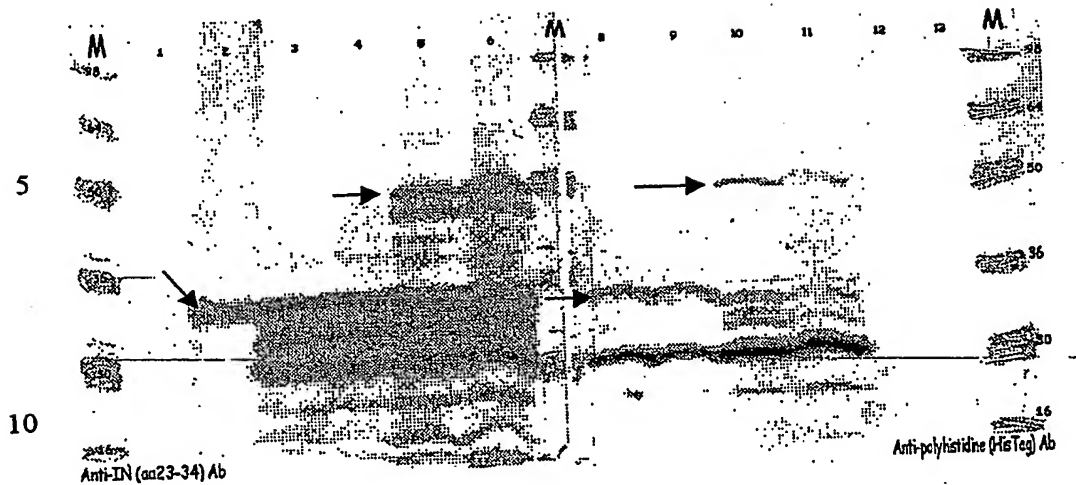
Figure 5.4.2-II: Immunoblot from pDcIN and pDIH expression in BL21-AI cells at 3 hours post induction. The left half of the blot was stained with anti-integrase antibody and the right half with anti-poly-histidine antibody. M: Protein molecular weight standard (SeeBlue® Pre-Stained Standard, Invitrogen); Lane 1: BL21-AI cell control; lane 2: positive control (pINSD.His 3h); lanes 3-4: pDcIN samples from parallel transformations (#1 and #2); lanes 5-6: pDIH samples from parallel transformations (#1 and #2); lane 8: M; lanes 8-9: same samples as in lanes 3-4; lanes 10-11: same samples as in lanes 5-6; lane 12: positive control (pINSD.His 3h; very weakly visible); lane 13: BL21-AI cell control. Staining with both antibodies shows numerous degradation products of the expressed proteins. Arrows show the expected sizes of the proteins.
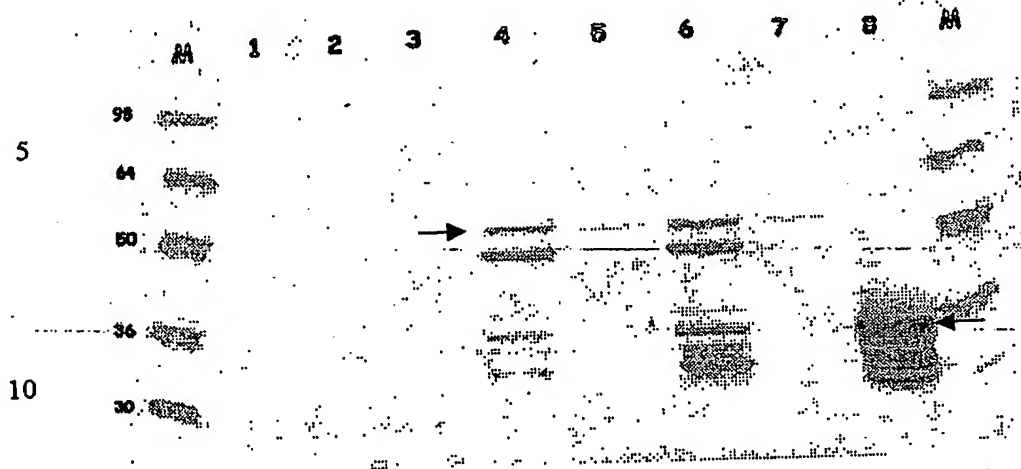
Figure 5.4.2-III: Immunoblot from pDIP expression in BL12-AI testing the impact of glucose in the growth media (=pDIPg). The blot was immunostained with the anti-IN antibody. M: Protein molecular weight standard (SeeBlue® Pre-Stained Standard, Invitrogen); Lane 1: BL21-AI cell control; lane 2: pDIP1 UI; lane 3: pDIPg UI; lane 4: pDIP1 1h; lane 5: pDIPg 1h; lane 6: pDIP1 2h; lane 7: pDIPg 2h; lane 8: positive control pINSD.His 1h. Black arrows show expected sizes of the proteins. The effect of glucose to protein degradation was unobservable but the effect on the level of protein expression was strong. The pDIP-protein expressed in absence of glucose is degraded like as noted previously with pDIH and pDcIN proteins.
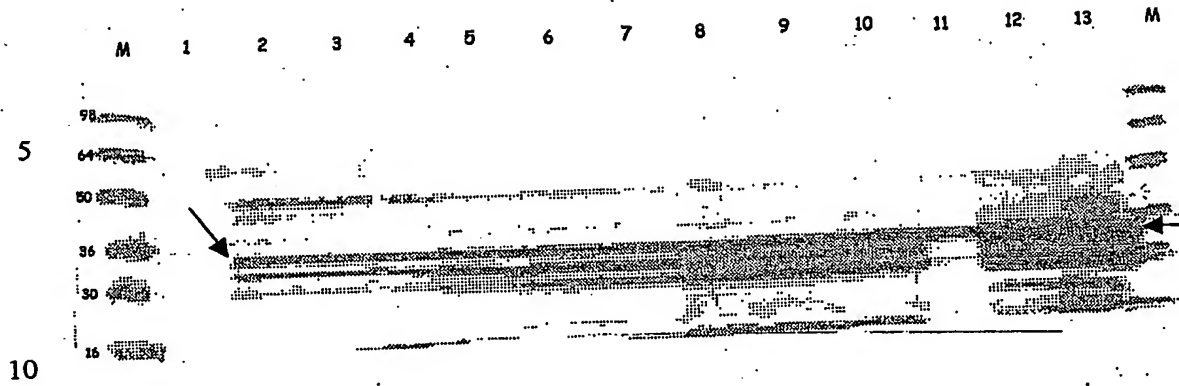
Figure 5.4.2-IV: DcIN protein expression in BL21-AI cells testing the impact of different inducer concentrations and lower incubation temperature (30°C) on protein expression levels and degradation. M: Protein molecular weight standard (SeeBlue® Pre-Stained Standard, Invitrogen); lane 1: BL21-AI cell control; lanes 2,3 and 4: uninduced (UI) DcIN expression samples (with quite strong basal expression levels) of 1mM, 0.1mM and 0.01mM IPTG induction, respectively; lanes 5,6 and 7: same samples 1h post induction; lanes 8,9 and 10: same samples 3h post induction; lanes 11, 12 and 13: positive control pINSD.His samples with 1mM IPTG induction, UI, 1h and 3h, respectively.

Figure 5.4.2-V: pDIH protein expression in BL21-AI cells testing the impact of different inducer concentrations and lower incubation temperature (30°C) on protein expression levels and degradation. M: Protein molecular weight standard (SeeBlue® Pre-Stained Standard, Invitrogen); lanes 1, 2 and 3: Uninduced (UI) pDIH samples 1mM, 0.1 mM and 0.01 mM IPTG induction, respectively; lanes 4,5 and 6: same samples 1h post induction; lanes 7, 8 and 9: same samples 3h post induction; lanes 10, 11 and 12: positive control sample pINSD.His with 0.1mM IPTG indu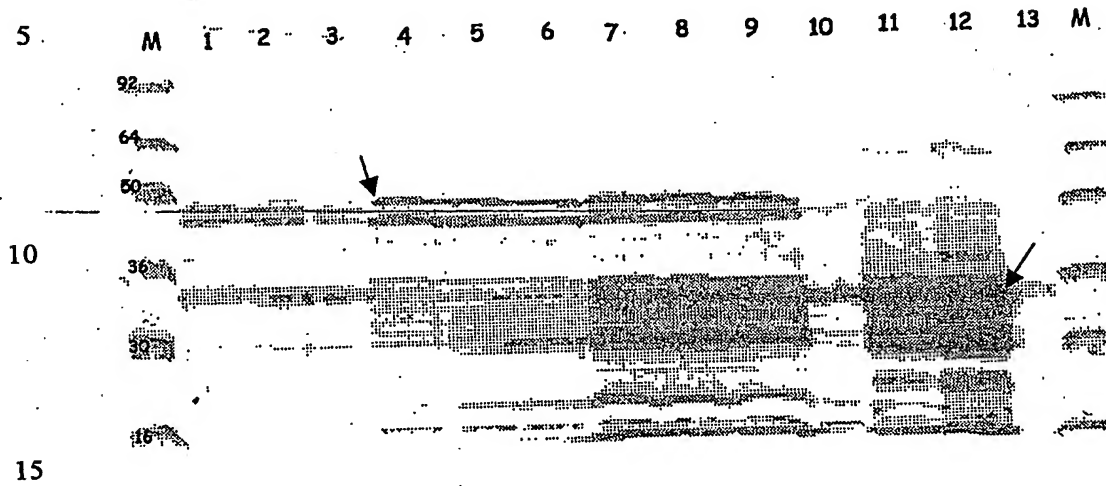ction, samples UI, 1h and 3h, respectively; lane 13: BL21-AI cell control. Black arrows show the expected sizes of the proteins.

6 DISCUSSION

The choice of the vector type for gene therapy applications will affect the persistence and stability of the introduced gene in the treated cells. Non-integrating vectors, such as adenoviruses, herpes-simplex viruses, baculoviruses and Rep-deleted

5 episomal rAAVs can mediate transgene expression for restricted periods of time. Sometimes the non-integrated vector DNA may however persist in the cell for even 9 months, as observed for rAAV (Pfeifer & Verma, 2001; Kootstra & Verma, 2003). In contrast, integrating vectors usually stabilise the persistence of the transgene in the host genome conferring long-term expression of the introduced gene. As noted earlier,

10 integration does not always ensure stable and constant expression of the transgene, as factors associated with chromatin structure and activity may affect the expression levels of transgenes and lead to vector silencing (Pannell & Ellis, 2001). Other problems associated with integrating vectors are the possibility of insertional mutagenesis and the inappropriate activation or inactivation of nearby genes by the integrated vector DNA

15 (Verma & Somia, 1997; Kootstra & Verma, 2003). The French study is a good example of a gene therapy trial ending up unexpectedly. Children suffering from X-linked severe combined immunodeficiency disease (SCID-X1) were treated by retrovirus vector based gene therapy (Cavazzana-Calvo et al., 2000). First, the treatment was able to provide clear clinical benefit by fully correcting the disease phenotype. Three years after the

20 reported success two of the cured children had developed a leukemia-like illness resulting from the integration of retrovirus genome near an oncogene (Marshall, 2003). Retroviral gene therapy has also caused leukemia in mice (Li et al., 2002). Both of these examples emphasise how important it is to develop integrating vectors capable of targeted transgene integration into benign loci of the human genome in order minimize

25 the risks associated with random transgene integration.

The aim of this study was to create a novel fusion protein IN-I-*Ppo*I for use in gene therapy applications. The new fusion protein was designed to be capable of catalysing targeted integration of retrovirus-like DNAs into a benign locus in the human genome. The aim also included the preparation of DNA constructs required for testing

30 the IN-I-*Ppo*I fusion proteins' activity *in vitro*. The homing endonuclease I-*Ppo*I

recognises and cleaves its target site present in the 28S rDNA of humans (Monnat et al., 1999). In conjunction with the HIV-1 IN it is designed to target IN-mediated integration of the therapeutic genes into its rDNA site in humans. Heterologous gene expression has previously been reported from the I-*Ppo*I site in the yeast rDNA (Lin & Vogt, 2000), which proves that functional proteins –possibly also therapeutic ones- can be expressed as pol I transcripts from the rDNA of eukaryotic cells.

6.1 In vitro tests for HIV-1 IN

The earliest *in vitro* assays studying the integration mechanisms of retroviruses used preintegration complexes (PICs) directly isolated from infected cells (Brown et al., 1987; Farnet & Haseltine, 1990; Lee & Coffin, 1990). The more recent assays use the recombinant purified IN protein of different retroviruses in conjunction with short synthetic oligonucleotide substrates that mimic the ends of the viral DNA molecule. In optimal conditions, these assays have shown that the purified integrase alone can catalyse the major steps of integration, namely the 3' end processing and the 3' end joining (strand transfer) reactions (Bushman & Craigie, 1991; Craigie et al., 1990; Katz et al., 1990).

The 3' end processing activity of the IN requires the purified protein and radioactively labelled double stranded oligonucleotides containing sequences derived from the ends of either the U5 or the U3 LTRs (Chow, 1997; Brown, 1997). Gel electrophoresis is conventionally used to resolve the shortened products from the substrates. The 3' end joining (strand transfer) reaction can be carried out with a similar composition, but the substrate is often "pre-processed", i.e. lacks the dinucleotides cleaved during the 3' modification step. The IN does not recognise special features of the integration target *in vitro*, so the target can either be the same molecule as the substrate or a circular plasmid DNA. The joining reaction can be assayed by the appearance of products that are longer in length than the input DNA in gel electrophoresis. The integration products may optionally be first phenol-chloroform extracted and ethanol precipitated, after which they can serve as templates for PCR (Chow, 1997). In addition to facilitating the detection of different sized products, PCR can also be used to facilitate the sequencing of the integration sites to reveal possible

integration hot spots and the characteristic 5 bp duplications created by HIV-1 DNA integration. The site of the joining is largely random for the wt HIV-1 IN and the lengths of the products may therefore vary greatly (Chow, 1997; Brown, 1997).

The concerted integration (3' end joining or strand transfer) of two viral DNA ends into a target DNA is harder to achieve *in vitro*. Often the *in vitro* assays described above result in the joining of one DNA end into one strand of the target DNA (half-site integration) (Chow, 1997; Brown, 1997). The concerted integration can more accurately be studied using isolated PICs that (in addition to IN) provide the accessory enzymes required for full-site integration (see below). A circular plasmid DNA is used as the integration target and the event can be assayed by the presence of a selectable marker in the recombinant product or by southern blotting. This assay better resembles integration of retroviruses *in vivo*, as the recombinant HIV-1 IN has been shown to require the virally encoded nucleocapsid (NC) protein (Carteau et al., 1999) or the host-derived HMG-I(Y) (Hindmarsh et al., 1999) to perform efficient full-site integration *in vitro*. Even though the consensus is that at least the viral NC is required for full-site integration *in vitro* (e.g. Brown, 1997; Carteau et al., 1999), Sinha et al. (2002) suggested that using recombinant IN alone it is possible to obtain full-site integration of two donor termini without any cellular or viral protein cofactors *in vitro*. They concluded that the key factor allowing recombinant wild-type IN to mediate full-site integration is the avoidance of high IN concentrations in its purification and in the integration assay. However, the reaction end-products that they accounted for full-site integration products involved the concerted insertion of two LTR ends (U5 and U3 LTRs) per target DNA, not discriminating whether the LTRs were from the same substrate molecule or from two separate ones. Such concerted integration products are not comparable with the full-site integration products of retroviral genomes that form the provirus.

Being aware of the possible end-results of the *in vitro* integration assays, this study was designed to provide the minimal components for the first step in testing the targeting abilities of a new fusion protein IN-I-*Ppo*I. Possible half-site integration products will reveal that the HIV-1 integrase retains its activity in the context of the chimeric recombinant fusion protein. Analysis of the integration sites on the target DNA plasmid will shed light on the targeting possibilities by the proteins fused to HIV-1

integrase. Also, the endonucleolytic activity of I-*Ppo*I or its mutant form (I-*Ppo*I-) H98A fused to IN can be assessed by a simple cleavage experiment using the I-*Ppo*I recognitions site plasmid pPPOsite. In case the *in vitro* integration experiment should result in full-site integration products, a marker gene (EGFP) is included in the

5 integration substrate to facilitate the screening of these events. Integration sites can be screened by PCR or RE digestion in either case, or radioactive labels can be incorporated to the integration substrate for faster observation by autoradiograms. An outline of the expected *in vitro* integration process is provided in figure 6.1-I.

(A)

| GTACTGGAAGGGCTAATTCACTGCATG |
| CATGACCTTCCCGATTAAGTGAC GTAC |

EGFP

| CATGCTGTGGAAAATCTCTAGCAGTAC |
| GTACGACACCTTTTAGAGATCGTCATG |

+

Donor DNA (2LTR-EGFP; 2650 bp)

Acceptor DNA (pPPOsite or pBSII; 3000bp)

● HIV-1 IN / IN-I-*Ppo*I / IN-H98A

(B)
Integration products:

Concerted or full-site integration   half-site integration products
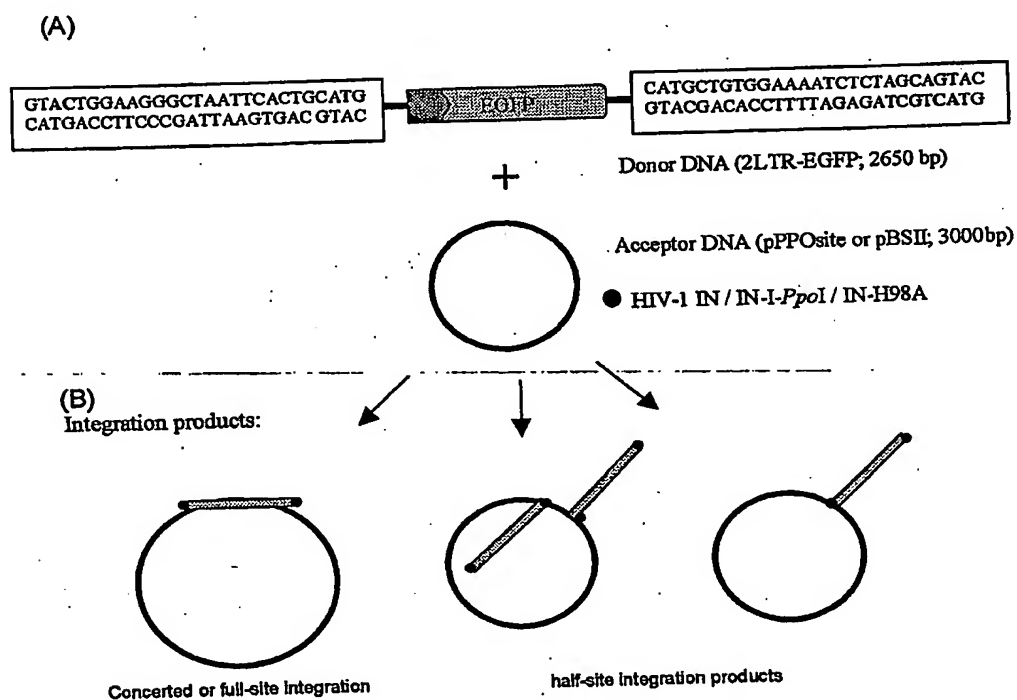
10 **Figure 6.1-I.** Reconstitution of integration *in vitro*. (A) A diagrammatic representation of a HIV-1 donor DNA substrate of ~2650 bps. The solid box represents an expression cassette for the EGFP (enhanced green fluorescent protein). (B) Diagrammatic representation of an in vitro integration reaction with an acceptor DNA (pPPOsite and as a control pBluescript II), integration donor DNA (substrate from pB2LTR+EGFP) and purified HIV-1 IN (cIN), IN-I-*Ppo*I or IN-

15 H98A proteins. Possible products include those that result from concerted integration of both donor DNA termini (left) and from non-concerted integration by two or more (middle) or one (right) donor DNA molecules via one-ended integration events.

6.1.1 Generating an integration target plasmid

To study the directed integration possibly achieved with the fusion proteins IN-I-*Ppo*I or IN-H98A, an integration target plasmid with the recognition site for I-*Ppo*I was created. The 15 bp sequence for the recognition site was adopted from previous studies using I-*Ppo*I (Mannino et al., 1999; Monnat et al., 1999; Ellison & Vogt, 1993; Argast et al., 1998). The recognition site was generated by hybridising two complementary oligonucleotides G515 and G517 (Appendix I, table I-4), after which the linker was cloned to the *Eco*RV site of pBluescript (Stratagene). Insertion of the linker was verified by both I-*Ppo*I (Promega) cleavage and by sequencing. pBluescript without the I-*Ppo*I recognition site was planned to serve as a control plasmid for the *in vitro* integration test to highlight the possible differences in the integration patterns of the wt HIV-1 IN (cIN in this study) and the proteins IN-I-*Ppo*I and IN-H98A.

6.1.2 Preparing the LTR-flanked integration substrate

The HIV-1 integrase, like other retroviral integrases, recognises special features at the ends of the viral DNA located in the U3 and U5 regions of the long terminal repeats (LTRs) (Brown, 1997). The LTR termini are the only viral sequences thought to be required in *cis* for recognition by the integration machinery of retroviruses. Short imperfect inverted repeats are present at the outer edges of the LTRs in both murine and avian retroviruses (reviewed by Reicin et al., 1995). Along with the subterminal CA located at the outermost positions 3 and 4 in retroviral DNA ends (positions 1 and 2 being the 3' end processed nucleotides; see figure 2.4.1-I), these sequences are both necessary and sufficient for correct proviral integration *in vitro* and *in vivo*. However, also sequences internal to the CA dinucleotide appear to be important for optimal IN activity (Brin & Leis, 2002a; Brin & Leis, 2002b; Brown, 1997). The terminal 15 bp of the HIV-1 LTRs have been shown to be crucial for correct 3' end processing and strand transfer reactions *in vitro* (Reicin et al., 1995; Brown, 1997). Longer substrates are used more efficiently than shorter ones by HIV-1 IN which indicates that binding interactions extend at least 14-21 bp inward from the viral DNA end. Brin and Leis (2002a) analysed the specific features of the HIV-1 LTRs and concluded that both the U3 and U5 LTR recognition sequences are required for IN-catalysed concerted DNA integration, even

though the U5 LTRs are more efficient substrates for IN processing *in vitro* (Bushman & Craigie, 1991; Sherman et al., 1992). The positions 17-20 of the IN recognition sequences are needed for a concerted DNA integration mechanism, but the HIV-1 IN tolerates considerable variation in both the U3 and U5 termini extending from the invariant subterminal CA dinucleotide (Brin & Leis, 2002b).

For this study the wild type 20 bp HIV-1 IN recognition sites of the viral LTRs were adopted from Brin and Leis (2002a) who had used a mini donor DNA substrate for concerted integration *in vitro*. As described previously, the deoxyoligonucleotides were designed in a way that the EGFP expression cassette could be cloned between the pBluescript contained 5' and 3' LTRs. In addition, the 5' and 3' LTRs were designed to contain a unique *Sca*I site that enables the digestion of the correctly blunt ended LTR-EGFP-LTR (2LTR-EGFP) integration substrate construct. A blunt ended integration substrate was created instead of a pre-processed one because it would allow detection of both the 3' end processing and the strand transfer reactions catalysed by IN. The EGFP marker gene was inserted between the LTRs to facilitate the recognition of full-site integration products that could perhaps be transformed into bacteria and selected in a biological assay (Brin & Leis, 2002a). The construct pBVboostFG+EGFP, from which the EGFP expression cassette was cloned, bears promoters for protein expression in insect cells, mammalian cells and bacteria (Laitinen et al., unpublished results). All these promoters were retained in the LTR-EGFP construct cloned into pBluescript to facilitate the possible switch of the host for the future biological assay of *in vitro* integrants.

6.1.3 Creation of expression constructs

The integrase of HIV-1 has previously been shown to retain its activity when fused to DNA binding proteins and to become targeted by the sequence specific protein to which it is fused (Bushman, 1994; Bushman, 1995; Goulaouic & Chow, 1996; Pavletich & Pabo, 1991; Bushman & Miller, 1997; Holmes-Son & Chow, 2000). In the present study, the gene for the HIV-1 integrase (AF029884, clone HXB2) was fused to the homing endonuclease I-*Ppo*I (M38131) and to its mutant form H98A. Also a single form of the wt HIV-1 IN was subcloned to provide a control for the future *in vitro* tests. First, the fusion of the IN to its fusion partners was achieved through subcloning the

cDNAs into pBluescript. Next, the cDNAs of IN-I-*Ppo*I, IN-H98A and cIN were transferred to pBVboostFG (Laitinen et al., unpublished results) which is a universal expression vector compatible with recombinational cloning based on the bacteriophage λ recombination system (Landy, 1989) of the Gateway™ cloning technology

5 (Invitrogen). The advantages of cloning the constructs IN-I-*Ppo*I, IN-H98A and cIN into this vector lie in the versatility of the vector driven expression of the cloned genes in various cells (Laitinen et al., unpublished results). Also, cloning the cDNAs into this vector is simple and has the advantage that baculoviral bacmids can easily be generated from the constructs residing in pBVboostFG (Laitinen et al., unpublished results).

10 Baculoviruses are arthropod viruses that have long been used in expressing recombinant genes in insect cells and more recently as mammalian cell gene-delivery vectors (Kost & Condreay, 2002; Hüser & Hofmann, 2003).

6.2 Protein expression and characterisation

6.2.1 Hosts for recombinant protein expression

15 Recombinant proteins can be overexpressed in various host cells. The most common and cost-effective expression host has for a long time been bacteria, of which special *E. coli* strains have been developed to optimise recombinant protein yields and convenience of protein production (Georgiou, 1996). The advantages associated with bacterial protein expression systems are the ease of protein production and cell growing,

20 time-saving and cost-effective expression protocols, versatility of bacterial phenotypes for different needs and usually high expression levels of most heterologous proteins. Prokaryotic expression systems are, however, deficient of the posttranslational protein modification systems provided by eukaryotic cells. When expressing a eukaryotic protein in a prokaryotic host, this factor may cause problems in protein yield, solubility

25 or activity if the expressed protein is incorrectly (or inadequately) folded (Georgiou, 1996). The production of proteins that accumulate in misfolded forms (usually inclusion bodies) is a common problem associated with bacterial protein expression, as is the proteolytic degradation of expressed proteins in *E. coli*. Eukaryotic expression hosts, such as insect or mammalian cell cultures, are more suitable for expression of proteins

30 that naturally become modified (e.g. glycosylated) when expressed in their natural host.

The baculovirus-based expression in insect cells has particularly the advantage of being a rapid and efficient eukaryotic translation environment in which active heterologous proteins can be produced (Kost & Condreay, 2002; Hüser & Hofmann, 2003; Gierse et al., 1993). Expression systems based on mammalian cell cultures may be more laborious but further ensure the correct environment for certain expressed proteins.

6.2.2 Expression of HIV-1 IN and IN-fusion proteins

Numerous studies have shown that the HIV-1 integrase can be expressed in *E. coli* cells and used in *in vitro* assays after partial or extensive purification of the enzyme (Sherman & Fyfe, 1990; Bushman, 1994; Engelman & Craigie, 1995; Goulaouic & Chow, 1996; Leh et al., 2000; Brin & Leis, 2002a and 2002b). IN has also been expressed in other systems, such as insect cell cultures by baculovirus-mediated expression (Bushman et al., 1990; Bischerour et al., 2003; Pruss et al., 1994). Often, the majority of the IN expressed either in *E.*coli or insect cells has been found in the insoluble fraction following centrifugation of expression cell lysates (e. g. Sherman & Fyfe, 1990; Bushman & Craigie, 1991). The low solubility of the IN protein is a known fact that also impedes its proper crystallisation and tertiary structure determination (Freed & Martin, 2001).

6.2.3 Degradation of cIN and IN-fusion proteins in bacterial cells

As it had been possible to obtain catalytically active bacterially expressed HIV-1z IN in previous studies (Sherman & Fyfe, 1990; Bushman, 1994; Engelman & Craigie, 1995; Goulaouic & Chow, 1996; Leh et al., 2000; Brin & Leis, 2002a and 2002b), the IN and its fusion constructs IN-I-*Ppo*I and IN-H98A were first produced in *E. coli*. Two bacterial strains were used to find a suitable expression host. Expression of the control (wt) integrase and the fusion construct IN-H98A proteins in both the bacterial strains *E.coli* BL21 (DE3) and BL21-AI resulted in high protein expression levels of both recombinant proteins, as observed by specific anti-integrase and anti-polyhistidine antibody staining of immunoblots. The expression of the fusion protein IN-I-*Ppo*I was hard to achieve and only slight amounts of the protein could be detected in the immunoblots. However strong the expression levels were for cIN and IN-H98A, significant degradation of the expressed proteins was detected in all experiments with

both the host strains (see figures 5.4.2-I – 5.4.2-V). Since all protein production experiments consisted of analytical scale expression without further purification it was not determined whether the majority of the proteins were contained in the soluble or the insoluble fraction after cell lysis. Expression samples were prepared for SDS-PAGE

5   analysis in a way to minimise protein degradation by proteases by immediate sample boiling and freezing (or vice versa). Still, all of the produced proteins were found to become degraded during bacterial expression. Production of HIV-1 IN in bacteria has also previously resulted in at least two forms of the protein (e.g. Sherman & Fyfe, 1990) but the extent of degradation observed in this study has not been reported before.

10   Protein degradation, or proteolysis, of recombinant proteins in *E. coli* is a common problem and may significantly affect the yields of heterologous proteins (Georgiou, 1996). The use of protease-deficient *E. coli* strains, as in this study, should prevent excessive proteolysis of the recombinant proteins. However, despite of using two different strains that both lack the proteases *lon* and *OmpT* capable of degrading

15   proteins during purification (Grodberg & Dunn, 1988; Studier & Moffat, 1986), protein degradation was observed in all experiments using the bacterial hosts in our study.

Optimisation of bacterial growth conditions may also prevent protein degradation during expression (Georgiou, 1996) but the changes tested in this study had no effects on observed proteolysis. Many proteins expressed in *E.coli* are known to form inclusion

20   bodies partly due to the host cell response for production of nonessential proteins (Allen et al., 1992). Formation of these tightly packaged structures may protect proteins from proteolysis in the cell and aid in protein purification (Georgiou, 1996). Although both HIV-1 IN and I-*Ppo*I are known to form the insoluble inclusion bodies in bacteria (e.g. Zheng et al., 1996; Flick et al., 1997; Mannino et al., 1999) this feature did not protect

25   the proteins from degradation in this study. As seen in the immunoblots derived from protein expression samples (see figure 5.4.2-II in Results), the IN fusion proteins and cIN tend to become cleaved in the C-terminal part of the IN protein, in the immediate N-terminus of IN-H98A samples (after 6x His residues and before aa 23 of IN, that is bound by the anti-IN-antibody; compare lanes 5 and 6 to 10 and 11 in figure 5.4.2-II)

30   and somewhere in the junction of the proteins IN and H98A. However, also correctly full-length proteins (cIN, IN-I-*Ppo*I and IN-H98A) were produced in the bacteria tested

(figures 5.5.2-I to 5.5.2-V). It is possible that the full-length proteins were insoluble as in inclusion bodies (and thus protected from proteolysis after all) and the degraded proteins were soluble and more prone to proteolysis. If this was the case, it would have been possible to separate the degraded proteins from the correct ones at least to some extent.

5 The expression host was however changed to insect cell culture before testing this because of the toxicity problems associated with production of IN-I-*Ppo*I in bacteria. Interestingly, almost no degradation of the proteins was detected as the expression host was switched to Sf9-insect cells. Further analysis of this observation is beyond the scope of this work but suggests that an *E. coli* host may not have been ideal for expression of

10 the particular proteins in this study. When expressed in insect cells, cIN was found to form inclusion bodies but reasonable amounts of the fusion constructs IN-I-*Ppo*I and IN-H98A were found in the soluble fraction. Neither fraction suffered from excessive proteolysis. It is hard to point out reasons for the unsuccessful expression of the proteins in *E.coli* since recombinant HIV-1 IN expression has previously been reported to work

15 in bacterial hosts (Sherman & Fyfe, 1990; Bushman, 1994; Engelman & Craigie, 1995; Goulaouic & Chow, 1996; Leh et al., 2000; Brin & Leis, 2002a and 2002b). The possibility of proteolysis occurring at some point of the SDS-PAGE sample preparation rather than inside bacterial cells cannot be fully excluded.

6.2.4 The toxicity of IN-I-*Ppo*I protein may reflect its biological
20 **activity**

As discussed above, expression of IN-I-*Ppo*I protein (from vector pDIP2) resulted in poor bacterial growth and very little amounts of expressed protein. Inhibition of bacterial growth by the expressed protein usually suggests that the gene product is toxic for the expression host (Georgiou, 1996). The inhibiting impact was not observed

25 with other expressed proteins, that is IN-H98A and cIN, which indicates that the toxic effect of IN-I-*Ppo*I must be associated with the I-*Ppo*I part of the fusion protein. The fusion protein consisting of IN and the endonucleolytically inactive I-*Ppo*I form (H98A) was not found to be toxic for the bacterial expression hosts. The homing endonuclease I-*Ppo*I has in fact been reported to cause cell death or reduce cell survival when expressed

30 in human cells (Monnat et al., 1999) and yeast (Muscarella & Vogt, 1993). Jurica &

Stoddard (1999) too reviewed that the over-expression of most HEs in common bacterial and yeast strains induces cell lysis and death. Toxicity of the I-*Ppo*I protein for human cells is likely to result from its endonucleolytic activity that causes efficient cleavage of its homing sites present in the rDNA.

5   The rDNA of the *E.coli* genome differs from the I-*Ppo*I recognition sequence in *P. polycephalum* at only 3 positions (Lowery et al., 1992), yet the expression of I-*Ppo*I has been facile and efficient in *E. coli* (Lowery et al., 1992; Mannino et al., 1999). These observations suggest that I-*Ppo*I has a low tolerance for degeneracy within its recognition sequence but are in sharp contrast with the growth inhibition observed in the present study. In the literature, I-*Ppo*I expression and activity in human cells has been associated with efficient cleavage of the homing site present in plasmids and in the rDNA of the host (Monnat et al., 1999). Double strand break repair (DSBR) of the cleaved DNAs was in some cases mutagenic and led to the loss of I-*Ppo*I recognition sites in the homing site plasmids after I-*Ppo*I cleavage and rejoining *in vivo*. Turning this point of view around, it could perhaps be possible that the genomes of the *E. coli* strains used in this study would contain a mutation to create new recognition and cleavage sites for I-*Ppo*I. This is however unlikely since the growth inhibition was observed with both the bacteria of different origins. A more likely explanation is that the fusion protein IN-I-*Ppo*I exhibits more sequence degeneracy than the wt I-*Ppo*I (perhaps due to unspecific DNA binding by IN), or has a novel enzymatic activity that is detrimental for bacteria. No growth inhibition was observed in insect cells that were tested for expression after experiments with bacteria. These results suggest that the novel fusion protein IN-I-*Ppo*I contains some enzymatic activity, perhaps enhanced DNA cleavage catalysis, and that the expression constructs themselves are functional. Furthermore, the good results obtained from expressing IN-I-*Ppo*I in insect cells overcome the problems associated with the production problems in bacteria.

6.3. The potential of integrating therapeutic genes into the hrDNA

The I-*Ppo*I itself is a good but rare example of a protein that becomes expressed as a RNA polymerase I transcript from the rDNA of its host (Lin & Vogt, 1998). When

expressed from the I-*Ppo*I-containing intron inserted in the yeast rDNA, the pol I-derived I-*Ppo*I mRNA was translated at approximately 3% of the efficiency of the pol II-derived RNA (Lin & Vogt, 2000). In the yeast rDNA, the α-fragment of β-galactosidase was inserted in place of the I-*Ppo*I ORF contained within the intron 3

5   (PpLSU3) of *P. polycephalum* (Lin & Vogt, 2000). The functional α-fragment was found to become expressed as a pol I transcript at similar levels as was shown for the I--*Ppo*I-endonuclease. Although expression levels were not high, the insertion and expression of the α-fragment in the rDNA of yeast demonstrated a new concept of functional protein expression from a pol I-derived RNA.

10   Transgene integration by HIV-1 IN can possibly be targeted to the rDNA of eukaryotes by the aid of I-*Ppo*I that recognises its target site embedded in the 28S rDNA gene. In case targeted integration into the rDNA locus is accomplished, expression of the transgene may in principle be carried out either by the RNA pol I constitutively expressing the rRNA genes, or by RNA pol II that may be recruited to the transgene site

15   by a strong promoter attached to the integerated transgene itself. The repeat organisers that act like insulators in the intergenic spacers of the rRNA genes (Robinett et al., 1997) may enhance the safety of the integrated transgene by separating its promoters and activity from the surrounding rRNA genes. The expression of a retrovirally integrated transgene in the rDNA will be tested in the future after the targeting capabilities of the

20   new fusion proteins IN-I-*Ppo*I and IN-H98A have been thoroughly studied.

6.4 Conclusions and future directions

The aim of this study was to create novel fusion proteins capable of catalysing targeted integration of retrovirus-like DNAs, and to supply them with the components necessary for testing the proteins' activity *in vitro*. Production of the proteins IN-I-*Ppo*I,

25   IN-H98A and cIN in *E. coli* cells was not successful as it resulted in proteolysis of all the proteins under study. In addition, low expression levels of IN-I-*Ppo*I were observed during the protein production experiments. This effect may have originated from the assumed toxicity of the I-*Ppo*I protein for *E. coli* cells. The excessive degradation of the proteins as well as their toxicity was however observed to occur within the bacterial

30   hosts only; later production of these proteins in insect cells was successful. The

purification of the proteins expressed in Sf9 cells needs to be further considered to obtain as native and active protein as possible. The DNA constructs pPPOsite (acceptor DNA) and pB2LTR+EGFP (donor DNA) created for the *in vitro* test are functional and will be used in future studies.

6.4.1 Future directions in protein production and activity tests

Bacterial expression hosts were found not to be optimal for expression of the proteins under study. In studies extending from this work, recombinant baculovirus genomes (from *Autographa californica* nuclear polyhedrosis virus; AcMNP) were created by transferring the cDNAs of cIN, IN-I-*Ppo*I and IN-H98A contained within the pBVboostFG to baculoviral genomes through a site specific transposition mechanism (Landy, 1989; Bac-to-Bac Baculovirus Expression system, Invitrogen; Luckow et al., 1993). Protein production was then driven in insect cells and it proved to be successful. Optimisation of the purification protocols has been accomplished for each protein and the *in vitro* testing of their activities is drawing closer. Aggregation of the insect cell expressed cIN proteins into inclusion bodies is posing new challenges to using the protein. cIN has to be purified under denaturing conditions and subsequently renatured to restore its conformation to the native and active form. There are many examples of HIV-1 IN renaturation protocols in the literature, and the *in vitro* activity of the protein does not change detectably if the protein is purified under native or denaturing conditions (e.g. Goulaouic & Chow, 1996). It also seems to be possible to purify soluble and correctly multimeric HIV-1 IN without detergents and in the presence of zinc (Leh et al., 2000) that becomes bound by the N-terminal HHCC-domain of HIV-1 IN. The presence of zinc has also previously been reported to enhance the multimerisation and catalytic activity of a soluble, mutant form of integrase (Zheng et al., 1996). Both Mg^{2+} and Zn^{2+} seem to be required to maintain an ordered, homogenous and enzymatically active tetramer of IN in solution (Leh et al., 2000). Also the homing endonuclease I-*Ppo*I contains a zinc binding motif in its His-Cys box (Jurica & Stoddard, 1999; Galburt et al., 2000; Chevalier & Stoddard, 2001). Zinc stabilises the loosely packaged tertiary structure of the homodimeric protein, which is important for its proper catalysis. The expression of soluble and active I-*Ppo*I in *E. coli* has been found to depend critically on

the presence of divalent metal ions in the growth medium; both Mg^{2+} and Zn^{2+} were essential for maintaining protein solubility in I-*Ppo*I purification (Flick et al., 1997). Knowledge of the importance of these metal ions for obtaining soluble and active protein will aid in future protein purification and renaturation.

6.4.2 From *in vitro* testing to cell experiments

Assuming that the *in vitro* activity of the proteins IN-I-*Ppo*I and IN-H98A will be successfully assayed, the proteins may next be tested in cell cultures. Initial plans have been made to introduce the purified proteins and an integration substrate into the nuclei of different cells to observe the integration patterns in cellular DNA. Also the expression of the proteins will be tested in human cells. It may be possible that the endonucleolytically active I-*Ppo*I will be toxic for eukaryotic cells as less than 40% of human cells (HT-1080 human fibrosarcoma cells) have survived from constitutive I-*Ppo*I expression lasting for 15-18 days (Monnat et al., 1999). I-*Ppo*I expressed in human cells is capable of cleaving approximately 10% its homing sites present in the 28S rDNA after 24 and 48 hours after transfection of the HE encoding plasmid (Monnat et al., 1999). If its activity is too high also when fused to HIV-1 IN, the endonucleolytically inactive form H98A should provide a means to target integration into the rDNA without homing endonuclease catalysed cleavage of the site. Also, it should be possible to design an I-*Ppo*I enzyme with reduced DNA cleaving activity or modified sequence specificity since the structure of the enzyme and the active site residues are well known (Jurica & Stoddard, 1999; Galburt et al., 2000; Chevalier & Stoddard, 2001). Ultimately, if the fusion proteins will be shown to target integration of substrate sequences resembling HIV-1 DNA, they may be incorporated into IN-mutated HIV-1 virions for example using the in *trans* approach of Bushman and Miller (1997). Also baculoviral hybrids may be tested for the same approach. The *in vivo* testing of the future vector derived from the work described here leads us back to the founding idea of this work: looking for a means to create a safe and efficient integrating vector for gene therapy.

LITERATURE

Allen, S. P., Polazzi, J. O., Gierse, J. K., Easton, A. M. (1992) Two novel heat shock genes encoding proteins produced in response to heterologous protein expression in Escherichia coli. *J Bacteriol.* 174:6938-6947.

Argast, G. M., Stephens, K. M., Emond, M. J. & Monnat, R. J. Jr (1998) I- *Ppo*I and I- *Cre*I homing site sequence degeneracy determined by random mutagenesis and sequential *in vitro* enrichment. *J. Mol. Biol.* 280:345-353.

Arnheim, N. & Erlich, H. (1992) Polymerase chain reaction strategy. *Annu. Rev. Biochem.* 61: 131-156.

Belfort, M. & Perlman, P. S. (1995). Mechanisms of intron mobility. *J. Biol. Chem.* 270: 30207-30240.

Belfort, M. & Roberts, R. J. (1997). Homing endonucleases: Keeping the house in order. *Nucleic Acids Res.* 25: 3379-3388.

Bischerour, J., Leh, H., Deprez, E., Brochon, J. C., Mouscadet, J. F. (2003) Disulfide-linked integrase oligomers involving C280 residues are formed in vitro and in vivo but are not essential for human immunodeficiency virus replication.*J Virol.* 77:135-141.

Blake, M.S., Johnston, K.H., Russell-Jones, G.J. and Gotschlich, E.C. (1984) A rapid, sensitive method for detection of alkaline phosphatase-conjugated anti-antibody on Western blots. *Anal.Biochem.* 136: 175-179.

Bouyac-Bertoia M, Dvorin JD, Fouchier RA, Jenkins Y, Meyer BE, Wu LI, Emerman M, Malim MH. (2001) HIV-1 infection requires a functional integrase NLS.*Mol Cell.* 7:1025-1035.

Brin, E. & Leis, J. (2002a) Changes in the mechanism of DNA integration *in vitro* induced by basesubstitutions in the HIV-1 U5 and U3 terminal sequences. *J. Biol. Chem.* 277:10938-10948.

Brin, E. & Leis, J. (2002b) HIV-1 integrase interaction with U3 and U5 terminal sequences *in vitro* defined using substrates with random sequences. *J. Biol. Chem.* 277:18357-18364.

Brin, E., Yi, J., Skalka, A. M., and Leis, J. (2000) Modeling the late steps in HIV-1 retroviral integrase-catalyzed DNA integration. *J Biol Chem* 275:39287-39295.

Brown, P. O. (1997) Integration. In: Coffin, J. M., Hughes, S. H. & Varmus, H. E (Eds.) *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (http://www.ncbi.nlm.nih.gov/books/bv.fcgi?tool=bookshelf&call=bv.View..ShowSection&sear chterm=HIV&rid=rv.figgrp.1557)

Brown, P. O., Bowerman, B., Varmus, H. E. & Bishop, J. M. (1987) Correct integration of retroviral DNA in vitro. *Cell* 3: 347-356.

Buchschacher, G. L. Jr, Wong-Staal, F. (2000) Development of lentiviral vectors for gene therapy for human diseases. *Blood* 95:2499-2504

Bukrinsky, M. I. & Haffar, O. K.(1999) HIV-1 nuclear import: in search of a leader. *Front. Biosci.* 4:772-781.

5 Burns, J. C., Friedmann, T., Driever, W., Burrascano, M. & Yee, J. K. (1993) Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cell. *Proc. Natl. Acad. Sci. USA* 90: 8033-8037.

10 Bushman, F. D. & Craigie, R. (1991) Activities of human immunodeficiency virus (HIV) integration protein *in vitro*: Specific cleavage and integration of HIV DNA. *Proc. Natl. Acad. Sci. USA* 88:1339-1343.

Bushman, F. D., Fujiwara, T. & Craigie, R. (1990) Retroviral DNA integration directedby HIV
15 integration protein in vitro. *Science* 4976:1555-1558.

Bushman, F.D. & Miller, M.D. (1997). Tethering human immunodeficiency virus type 1 preintegration complexes to target DNA promotes integration at nearby sites. *J. Virol.* 71: 458-464.
20
Bushman, F.D. (1994). Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. *Proc. Natl. Acad. Sci. USA* 91: 9233-9237.

Bushman, F.D. (1995). Targeting retroviral integration. *Science*: 267: 1443-1444.
25
Carteau S., Gorelick, R. J., Bushman, F. D. (1999). Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: Stimulation by the viral nucleocapsid protein. *J. Virol* 73:6670-6679.

30 Cavazzana-Calvo, M., Hacein-Bey, S., de Saint Basile, G., Gross, F., Yvon, E., Nusbaum, P., Selz, F., Hue, C., Certain, S., Casanova, J.L., Bousso, P., Deist, F.L., Fischer, A. (2000) Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science* 288: 669-72.

Chen, H., Engelman, A. (1998) The barrier-to-autointegration protein is a host factor for HIV
35 type 1 integration. *Proc Natl Acad Sci U S A.*95:15270-15274

Chevalier, B. S. & Stoddard, B. L. (2001). Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.*, 29: 3757-3774.

40 Chow, S. A. (1997) In vitro assays for activities of retroviral integrase.*Methods.* 12:306-317.

Cooper, G. M. (2000) In: The cell: a molecular approach, 2nd ed. Sinauer Associates, Inc Sunderland (MA) (http://www.ncbi.nlm.nih.gov/books/bv.fcgi?tool=bookshelf&call= bv.View..ShowSection&searchterm=rRNA&rid=cooper.section.1362#1363)
45
Costantini L. C., Jacoby, D. R., Wang, S., Fraefel, C., Breakefield, X. O. & Isacson, O. (1999) Gene transfer to the nigrostriatal system by hybrid herpes simplex virus / adeno- associated virus vector. *Hum Gene Ther* 10:2481-2494.
Craigie, R. (2001) Minireview: HIV integrase, a brief overview from chemistry to therapeutics.
50 *TheJournal of Biological Chemistry* 276:23213-23216.

Craigie, R., Fujiwara, T. & Bushman, F. (1990) The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro. *Cell* 62:829-837

Craigie, R., Hickman, A. B., Engelman, A. (1995) In: HIV Volume 2: A Practical Approach. Ed
5 J. Karn Oxford Univ Press, p. 53.

Dalgaard, J. Z., Garrett, R.A. & Belfort, M. (1993). A site-specific endonuclease endoded by a typical archaeal intron. *Proc. Natl. Acad. Sci. USA* 90:5414-5417.

10 Depienne C, Mousnier A, Leh H, Le Rouzic E, Dormont D, Benichou S, Dargemont C. (2001) Characterization of the nuclear import pathway for HIV-1 integrase. *J Biol Chem.* 276:18102-18107.

Dong J. Y., Fan P. D., Frizzell R. A. (1996) Quantitative analysis of the packaging capacity of
15 recombinant adeno-associated virus *Hum Gene Ther* 7:2101-2112

Dyer, M. R. & Herrling, P. L. (2000). Progress and potential for gene-based medicines. *Molecular Therapy* 1:213-224.

20 Dykxhoorn, D. M, Novina, C. D., Sharp, P. A. (2003) Killing the messenger: short RNAs that silence gene expression. *Nat Rev Mol Cell Biol.* 4:457-467.

Ellison, E. L. & Vogt, V. M. (1993) Interaction of the intron-encoded mobility endonuclease I-PpoI with its target site. *Mol. Cell. Biol.* 13:7531-7539.
25
Emi N., Friedmann T., Yee J. K. (1991) Pseudotype formation of murine leukemia virus with the G protein of vesicular stomatitis virus. *J Virol.* 65:1202-1207.

Engelman, A. & Craigie, R. (1992) Identification of conserved amino acid residues critical for
30 human immunodeficiency virus type 1 integrase function in vitro. *J Virol* 66:6361-6369.

Engelman, A. & Craigie, R. C. (1995) Efficient magnesium-dependent human immunodeficiency virus type 1 integrase activity. *J. Virol.* 69:5908-5911.

35 Engelman, A., K. Mizuuchi, and R. Craigie. (1991). HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transger. *Cell* 67: 1211-1221.

Farnet, C. M. & Bushman, F. D. (1997) HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegratin complexes in vitro. *Cell* 88: 483-492.
40
Farnet, C. M. & Haseltine, W. A. (1990) Integration of human immunodeficiency virus type 1 *in vitro Proc. Natl. Acac. Sci. USA* 87:4164-4168.

Flick, K. E., Jurica, M.S., Monnat Jr., R. J. & Stoddard, B.L. (1998). DNA binding and cleavage
45 by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature* 394: 96-101.

Flick, K. E., McHugh, D., Heath, J. D., Stephens, K. M., Monnat R. J. Jr & Stoddard, B. L. (1997) Crystallisation and preliminary X-ray studies of I-PpoI: a nuclear, intron-encoded homing enconuclease form Physarum polycephalum. *Protein Sci.* 6: 2677-2680.
50

Frankel, A. D. & Young, J. A. T (1998) HIV-1: Fifteen proteins and an RNA. *Annu. Rev. Biochem.* 67:1-25.

Freed, E. O. & Martin, M. A. (2001) HIVs and their replication. In: Knipe and Howley (Eds.) 5 *Fields Virology* 4th edn. pp 1971-2401. Philadelphia: Lippincott Williams & Wilkins.

Galburt, E. A., Chadsey, M. S., Jurica, M. S., Chevalier, B. S., Erho, D., Tang, W., Monnat Jr., R. J. & Stoddard, B. L. (2000). Conformational changes and cleavage by the homing endonuclease I-PpoI: a critical role for a leucine residue in the active site. *J. Mol. Biol.* 300: 877-10 887.

Gallay P, Hope T, Chin D, Trono D. (1997) HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway.*Proc Natl Acad Sci U S A.* 94:9825-9830.

15

Georgiou, G. (1996) Expression of proteins in bacteria. In: *Protein Engineering: Principles and Practice* pp. 101-127.Edited by Jeffrey L. Cleland and Charles S. Craik. Viley-Liss, Inc.

Gierse J. K., Luckow, V. A., Askonas, L. J., Duffin, K. L., Aykent, S., Bild, G. S., Rodi, C. P., 20 Sullivan, P. M., Bourner, M. J., Kimack, N. M., et al. (1993) High-level expression and purification of human leukotriene A4 hydrolase from insect cells infected with a baculovirus vector *Protein Expr Purif.*4: 358-366.

Gimble, F. (2001) Degeneration of a homing endonuclease and its target sequence in wild yeast 25 strain. *Nucleic Acids Research* 29:4215-4223.

Gimble, F. S.& Thorner, J. (1992) Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae. Nature* 357:301-306

30 Goff, P. G. (2001) *Retroviridae*: The retroviruses and their replication. In: Knipe and Howley (Eds.) *Fields Virology* 4th edn. pp 1871-1939.Philadelphia: Lippincott Williams & Wilkins.

Goulaouic, H., and Chow, S.A. (1996). Directed integration of viral DNA mediated by fusion proteins consisting of human immunodeficiency virus type 1 integrase and Escherichia coli 35 LexA protein. *J. Virol.* 70: 37-46.

Gould, D. J. & Favorov, P. (2003). Vectors for the tratement of autoimmune diseases.*Gene Therapy* 10:912-927

40 Graham, I. R. & Dickson, G. (2002) Gene repair and mutagenesis mediated by chimeric RNA-DNA oligonucleotides: chimeraplasty for gene therapy and conversion of single nucleotide polymorphisms (SNPs) *Biochim Biophys Acta.* 1587:1-6.

Grodberg, J.& Dunn, J. J. (1988) ompT encodes the Escherichia coli outer membrane protease 45 that cleaves T7 RNA polymerase during purification. *J Bacteriol.* 170:1245-1253.

Guo, H., Zimmerly, S., Perlman, P. S. & Lambowitz, A. M. (1997) Group II endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. *The EMBO Journal* 16: 6835-6848.

50

Haffar OK, Popov S, Dubrovsky L, Agostini I, Tang H, Pushkarsky T, Nadler SG, Bukrinsky M. (2000) Two nuclear localization signals in the HIV-1 matrix protein regulate nuclear import of the HIV-1 pre-integration complex. *J Mol Biol.* 299:359-368.

5   Hindmarsh, P., Ridky, T., Reeves, R., Andrake, M., Skalka, A. M., and Leis, J. (1999) HMG protein family members stimulate human immunodeficiency virus type 1 and avian sarcoma virus concerted DNA integration in vitro. *J. Virol.* 73: 2994-3003

10   Holmes-Son, M. & Chow, S.A. (2000) Integrase –lexA fusion proteins incorporated into human immunodeficiency virus type 1 that contains a catalytically inactive integrase gene are functional to mediate integration. *J. Virol* 24:11548-11556.

Holmes-Son, M. & Chow, S.A. (2002) Correct integration mediated by integrase-lexA fusion proteins incorporated into HIV-1. *Mol. Therapy* 5:360-370

15

Hong, T., Murphy, E., Groarke, J., Drilca, K. (1993) Human immunodeficiency virus type 1 DNA integration: fine structure target analysis using synthetic oligonucleotides.*J Virol* 67:1127-1131http://www.wiley.co.uk/genetherapy/clinical/

20   Hu, W.S. & Pathak, V.K. (2000) Design of retroviral vectors and helper cells for gene therapy. *Pharmacol. Rev.* 52: 493-511.

Hunter, E. (1997) Viral Entry and Receptors. (Chapter 3) In: Coffin, J. M., Hughes, S. H. & Varmus, H. E (Eds.) *Retroviruses.* Cold Spring Harbor Laboratory Press, Cold Spring

25

Harbor,N.Y.(http://www.ncbi.nlm.nih.gov/books/bv.fcgi?call=bv.View..ShowSection&rid=rv.chapter.571)

Hüser A, Hofmann C. (2003) Baculovirus vectors: novel Mammalian cell gene-delivery vehicles
30   and their applications. *Am J Pharmacogenomics* 3: 53-63

Johansen, S., Embley, T. M., Willassen, N. P. (1993). A family of nuclear homing endonucleases. *Nuceicl. Acids Res.* 21:4405-4411.

35   Jurica, M. S. & Stoddard, B. L. (1999). Homing endonucleases: structure, function and evolution. *CMLS, Cell. Mol. Life Sci.* 55:1304-1326.

Kalpana, G., V., Marmon, S., Wang, W., et al. (1994) Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science*; 266:2002-2006

40

Katz, R. A., Merkel, G., Kulkosky, J., Leis, J. & Skalka, A. M. (1990) The avian retroviral IN protein is both necessary and sufficient for integrative recombination in vitro. *Cell* 63:87-95

Katz, R.A. & Skalka A. M. (1994). The retroviral enzymes. *Annu. Rev. Biochem.* 63:133-173.

45

Kay, M. A., Glorioso, J. C., Naldini, L. (2001) Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics. *Nat Med. Jan*;7:33-40.
Kearns, W. G., Afione, S. A., Fulmer, S. B., Pang, M. C., Erikson, D., Egan, M., Landrum, M. J., Flotte, T. R., Cutting, G. R. (1996) Recombinant adeno-associated virus (AAV-CFTR)
50   vectors do not integrate in a site-specific fashion in an immortalized epithelial cell line. *Gene Ther.* 3:748-755.

Kootstra, N. A. & Verma, I. M. (2003) Gene Therapy with viral vectors. *Annu. Rev. Pharmacol. Toxicol.* 43:413-439.

5   Kost, T. A., Condreay, J. P. (2002) Recombinant baculoviruses as mammalian cell gene-delivery vectors. *Trends Biotechnol* 20:173-180.

Laitinen, O. H., Airenne, K. J., Peltomaa, E., Hytönen, V. P., Mähönen, A. J., Mäkelä, K. A., Kulomaa, M. S., Ylä-Herttuala, S. Manuscript: A tetra-promoter vector, pBVboostFG, for the

10·   simultaneous screening of ibraries in bacteria, insect and mammalian cells.

Landy, A. (1989) Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu Rev Biochem* 58:913-949.

15   Lee, N. (1980) Molecular Aspects of *ara* Regulation. In *The Operon*, J. H. Miller and W. S. Reznikoff, eds. Cold Spring Harbor, N.Y. :Cold Spring Harbor Laboratory, pp. 389-410.

Lee, N., Francklyn, C., and Hamilton, E. P. (1987). Arabinose-Induced Binding of AraC Protein to *ara*I2 Activates the *ara*BAD Operon Promoter. *Proc. Natl. Acad. Sci. USA* 84: 8814-8818.

20
Lee, Y. M. & Coffin, J. M. (1990) Efficient autointegration of avian retrovirus DNA in vitro. *J. Virol.* 12:5958-65.

Leh, H., Brodin, P., Bischerour, J., Deprez, E., Tauc, P., Brochon, J.-C., LeCam, E., Coulaud,

25   D., Auclair, C. & Mouscadet, J.-F. (2000) Determinants of Mg2+ activities of recombinant human immunodeficiency virus type 1 integrase. *Biochemistry* 39:9285-9294.

Lewin, B. (2000) Genes VII, Oxford University Press Inc. New York

30   Li, Z., Dullmann, J., Schiedlmeier, B., Schmidt, M., von Kalle, C., Meyer, J., Forster, M., Stocking, C., Wahlers, A., Frank, O., Ostertag, W., Kuhlcke, K., Eckert, H.G., Fehse, B., Baum, C. (2002) Murine leukemia induced by retroviral gene marking.

Lin, J. & Vogt, V. M. (1998). I-PpoI, the endonuclease encoded by the group I intron PpLSU3,

35   is expressed from an RNA polymerase I transcript. *Mol. Cell. Biol.* 18: 5809-5817.

Lin, J.& Vogt, V. M. (2000) Functional alpha-fragment of beta-galactosidase can be expressed from the mobile group I intron PpLSU3 embedded in yeast pre-ribosomal RNA derived from the chromosomal rDNA locus. *Nucleic Acids Res.* 28:1428-1438.

40
Loizos, N., Tillier, E. R., Belfort, M. (1994) Evolution of mobile group I introns: recognition of intron sequences by an intron-encoded endonuclease. *Proc Natl Acad Sci U S A.* 91: 11983-11987

45   Lowery, R., Hung, L., Knoche, K. & Bandziulis, R. (1992) Properties of I-PpoI: A rare-cutting intron-encoded endonuclease. *Promega Notes Magazine* 38:8
(http://www.promega.com/pnotes/38/38_08/38_08.htm)
Luckow, V. A., Lee, S. C., Barry, G. F., Olins, P. O. (1993) Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a

50   baculovirus genome propagated in Escherichia coli. *J Virol.* 67:4566-4579.

68

Mannino, S. J., Jenkins, C. L. & Raines, R. T. (1999) Chemical mechanism of DNA cleavage by the homing endonuclease I-PpoI. *Biochemistry* 38: 16178-16186.

Marshall, E. (2003) News: Second child in french trial is found to have leukaemia. *Science* 299:320Miao C. H., Snyder R. O., Schowalter D. B., Patijn G. A., Donahue B, Winther B., Kay M. A. (1998) The kinetics of rAAV integration in the liver. *Nat Genet.* 19:13-15.

Michelsen, B. K. (1995) Transformation of Escherichia coli increases 260-fold upon inactivation of T4 DNA ligase. *Anal Biochem.* 225:172-174.

Miller, A. D. (1997) Development and Applications of Retroviral Vectors (Chapter 9) In: Coffin, J. M., Hughes, S. H. & Varmus, H. E (Eds.) *Retroviruses.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Miller, D. G., Rutledge, E. A. & Russell, D. W. (2002). Chromosomal effects of adeno-associated virus vector integration. *Nature Genet.* 30: 147-148.

Miller, D.G., Adam, M.A. & Miller, A.D. (1990) Gene transfer by retrovirus vectors occurs only in cells that are actively replicating at the time of infection. *Mol. Cell Biol.* 10:4239-4242.

Miller, M.D., Bor, Y.C., Bushman, F. (1995). Target DNA capture by HIV-1 integration complexes. *Curr. Biol.* 5: 1047-1056.

Monnat Jr. R. J., Hackmann, A. F. M. & Cantrell, M. A. (1999) Generation of highly site-specific DNA double strand breaks in human cells by the homing endonucleases I-PpoI and I-CreI. *Biochem. Biophys. Res. Communications* 255: 88-93.

Murphy, S. J., Chong, H., Bell, S., Diaz, R. M. & Vile, R. G. (2002) Novel integrating adenoviral/ retroviral hybrid vector for gene therapy. *Hum. Gene Ther.* 13:745-760.

Muscarella, D. E & Vogt, V. M. (1993) A mobile group I intron from Physarum polycephalum can insert itself and induce point mutations in the nuclear ribosomal DNA of saccharomyces cerevisiae. *Mol. Cell. Biol.* 13:1023-1033.

Muscarella, D. E. & Vogt, V. M. (1989). A mobile group I intron in the nuclear rDNA of Physarum polycephalum. *Cell* 56: 443-454.

Muscarella, D. E., Ellison, E. L., Ruoff, B. M., Vogt, V. M. (1990). Characterization of I- *Ppo*I, an intron-encoded endonuclease that mediates homing of a group I intron in the ribosomal DNA of *Physarum polycephalum. Mol. Cell. Biol.* 10: 3386-3396.

Nakai, H., Montini, E., Fuess, S., Storm, T. A., Grompe, M., Kay, M. A. (2003) AAV serotype 2 vectors preferentially integrate into active genes in mice. *Nat Genet.* 34:297-302.

Nakai, H., Yant, S. R., Storm, T. A., Fuess, S., Meuse, L., Kay, M. A. (2001). Extrachromosomal recombinant adeno-associated virus vector genomes are primarily responsible for stable liver transduction *in vivo. J. Virol.* 75:6969-6976.

Olivares, E. C., Hollis, R.P, Chalberg, T. W., Meuse, L., Kay, M. A. & Calos, M. P. (2002) Site specific genomic integration produces therapeutic factor IX in mice. *Nature Biotechnology* 20: 1124-1128.

69

Ortiz-Urda, S., Thyagarajan, B., Keene, D. R., Lin, Q., Fang, M., Calos, M. P. & Khavari, P. A. (2002) Stable nonviral genetic correction of inherited human skin disease. *Nature Medicine* 8: 1166-1170.

5

Palombo, F., Monciotti, A., Recchia, A., Cortese, R., Ciliberto, G., La Monica, N. (1998) Site-specific integration in mammalian cells mediated by a new hybrid baculovirus-adeno-associated virus vector. *J Virol.*72:5025-34.

10   Pannell, D. & Ellis, J. (2001) Silencing of gene expression: implications for design of retrovirus vectors. *Rev. Med. Virol.* 11:205-217.

Pavletich, N. P., and Pabo, C.O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science* 252: 809.

15

Pfeifer, A. & Verma, I.M. (2001) Virus vectors and their applications. In: Knipe and Howley (Eds.) *Fields Virology* 4th edn. Philadelphia: Lippincott Williams & Wilkins.

Pruss, D., Bushman, F. D & Wolffe, A. P. (1994) Human immunodeficiency virus integrase
20   directs integration to sites of severe DNA distortion within nucleosome core. *Proc. Natl. Acad. Sci. USA* 91:5913-5917.

Recchia, A., Parks, R. J., Lamartina, S. Toniatti, C., Pieroni, L., Palombo, F., Ciliberto, G., Graham, F.L., Cortese, R., La Monica, N., Colloca, S. (1999) Site-specific integration mediated
25   by a hybrid adenovirus / adeno-associated virus vector. *Proc Natl Acad Sci USA* 96:2615-2620.

Reicin, A. S., Kalpana, G., Paik, S., Marmon, S. & Goff, S. (1995) Sequences in the human immunodeficiency virus type 1 U3 region required for in vivo and in vitro integration. *J. Virol.* 69:5904-5907.

30

Rice, P., Craigie, R., Davies, D. R. (1996). Retroviral integrases and their cousins. *Curr Opin Struct Biol.* 6: 76-83

Robinett, C. C., O'Connor, A. & Dunaway, M. (1997) The repeat organizer, a specialized
35   insulator element within the intergenic spacer of the Xenopus rRNA genes *Mol. Cell. Biol.* 17: 2866-2875

Roe, T., Reynolds, T.C., Yu, G. & Brown, P.O. (1993) Integration of murine leukemia virus DNA depends on mitosis. *EMBO J.* 12: 2099-2108.

40

Ruoff, B., Johansen, S. & Vogt, V. M. (1992) Characterization of the self-splicing products of a mobile intron from the nuclear rDNA of Physarum polycephalum.*Nucleic Acids Res.* 20:5899-5906

45   Sambrook, Russel. (2001). Molecular cloning. A laboratory manual third edition. Cold spring harbor laboratory press. Cold spring harbor, NY.
Satoh W, Hirai Y, Tamayose K, Shimada T. (2000) Site-specific integration of an adeno-associated virus vector plasmid mediated by regulated expression of rep based on Cre-loxP recombination. *J Virol*74:10631-10638

50

Schröder, A.R.W et al. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 110:521-529

Shagen, F. H. E., Rademaker, H. J., Fallaux, F. J., Hoeben R. C. (2000) Editorial: Insertion vectors for gene therapy. *Gene Therapy* 7: 271-272.

Sherman, M. P., de Noronha, C. M., Heusch, M. I., Greene, S., Greene, W. C. (2001) Nucleocytoplasmic shuttling by human immunodeficiency virus type 1 Vpr. *J Virol.* 75:1522-1532

Sherman, P. A., Dickson, M. L. & Fyfe, J. A. (1992) Human immunodeficiency virus type 1 integration protein: DNA sequence requirements for cleaving and joining reactions. *J. Virol.* 66:5393-601.

Sherman, P.A & Fyfe, J. A. (1990) Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity. *Proc. Natl. Acad. Sci. USA* 87:5119-5123.

Sinha, S., Pursley, M. H. & Grandgenett, D. P. (2002) Efficient concerted integration by recombinant human immunodeficiency virus type 1 integrase without cellular or viral cofactors. *J. Virol.* 76:3105-3113.

Sioud, M.& Drilca, K. (1991) Prevention of human immunodeficiency virus type 1 integrase expression in Escherichia coli by a ribozyme. *Proc Natl Acad Sci USA* 88:7303-7307

Studier, F. W. & Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol.* 189:113-130

Thomas, C. E., Ehrhardt, A., Kay, M. A. (2003) Progress and problems with the use of viral vectors for gene therapy. *Nat Rev Genet.* 4:346-358 hyagarajan, B., Olivares, E.C., Hollis, R.P., Ginsburg, D. S. & Calos, M. P. (2001). Site specific genomic integration in mammalian cells mediated by phage φC31 integrase. *Mol. Cell. Biol.* 21: 3926-3934.

Turner, B. G. & Summers, M. F. (1999) Structural biology of HIV. *J Mol Biol.* 285:1-32.
Ueno T, Matsumura H, Tanaka K, Iwasaki T, Ueno M, Fujinaga K, Asada K, Kato I. (2000) Site-specific integration of a transgene mediated by a hybrid adenovirus/adeno-associated virus vector using the Cre/loxP-expression-switching system. *Biochem Biophys Res Commun.* 273:473-478

Verma, I. M. & Somia, N. (1997). Gene therapy – promises, problems and prospects. *Nature* 389: 239-242.

Vigdal, T. J., Kaufman, C. D., Izsvák, Z., Voytas, D. F. & Ivics, Z. (2002) Common physical properties of DNA affecting target site selection of *Sleeping Beauty* and other Tc1/*mariner* transposable elements. *J. Mol. Biol.* 323:441-452.

Vink, C., Groenink, M., Elgerma, Y., Fouchier, R. A. M., Tersmette, M., and Plasterk, R. H. (1990) Analysis of the junctions between human immunodeficiency virus type 1 proviral DNA and human DNA. *J. Virol.* 64:5626-5627

Vogt, V. M. (1997) Retroviral Virions and Genomes (Chapter 2) In: Coffin, J. M., Hughes, S. H. & Varmus, H. E (Eds.) *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

5   Weitzman, M. D., Kyostio, S. R., Kotin, R. M. & Owens, R. A. (1994) Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integeration site in human DNA. *Proc. Natl Acad Sci USA* **91**:5808-5812

Wiesendanger, B., Lucchini, R., Koller, T., Sogo, J. M. (1994) Replication fork barriers in the
10   Xenopus rDNA. *Nucleic Acids Res.* **22**:5038-46.

Wittmayer, P. K. & Raines, R. (1996) Substrate binding and turnover by the highly specific I-*Ppo*I endonuclease. *Biochemistry* **35**: 1076-1083.

15   Wittmayer, P. K., McKenzie, J. L. & Raines, R. T. (1998) Degenerate DNA recognition by I-PpoI endonuclease. *Gene* **206**: 11-21.

Yant, S. R., Ehrhardt, A., Mikkelsen J. G., Meuse, L., Pham, T., Kay, M. A. (2002). Transposition from a gutless adeno-transposon vector stabilizes transgene expression *in vivo*.
20   *Nature Biotechnol.* **20**:999-1005.

Yant, S. R., Meuse, L., Chiu, W., Ivics, Z. Izsvak, Z. & Kay, M. A. (2000). Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. *Nature Genetics* **25**:35-41.
25
Zheng, R., Jenkins, T. M. & Craigie, R. (1996) Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc. Natl. Acad. Sci. USA* **93**:13659-13664.

30   Zimmerly, S., Moran J.V., Perlman, P. S. & Lambowitz, A. M. (1999) Group II intron reverse transcriptase in yeast mitochondria. Stabilisation and regulation of reverse transcriptase activity by the intron RNA. *J. Mol. Biol.* **289**: 473-490.

APPENDICES:

APPENDIX I

5   Table I-1 Primers used in insert-PCR (4.1.1.1). The sequences are presented in the 5' → 3' direction. RE-sites introduced in the primers are presented as underlined sequences and explained in the table. Start- and stop codons are bolded.

Primer name	Sequence	RE-site	Template& PCR product
F992	CC<u>TTAATTAA</u>**ATG**TTTTTAGATGGA ATAGAT	PacI	PLJS10, IN
3'IN	GC<u>TCTAGA</u>ATCCTCATCCTGTCTACT	XbaI	--''--, IN, cIN
3'cIN	TAT<u>GGCCTCTCAGGCC</u>**ATTATTAAT** CCTCATCCTGTCTACT	SfiI	--''--, cIN
G7	ATTCACC<u>ACTAGT</u>GCTCCAAAAAAA AAGCGC	SpeI	pCNPpo6: I-PpoI pCNPpo6h98a: H98A
F987	TAT<u>GGCCTCTCAGGCC</u>**ATTATTATA** CCACAAAGTGACTGCC	SfiI	----''----

10

Table I-2 Primers used in GATEWAY-PCR (4.1.7.1). The sequences are presented in the 5' → 3' direction. Two stop codons in GW 3'Ppo- and GW 3'cIN primers, as well as a start codon in the GW 5'IN HT, are marked in bolded. The six histidine encoding codons in GW 5'IN HT are underlined:

15

Primer name	Sequence
G238 (GW 3'Ppo)	GGGGACCACTTTGTACAAGAAAGCTGGGTTATGGCC TCTCAGGCCATTATTATACCACAAAGTGACTGCC
G402 (GW 3' cIN)	GGGGACCACTTTGTACAAGAAAGCTGGG**TATTATTA** ATCCTCATCCTGTCTACT
G445 (GW 5'IN HT)	GGGACAAGTTTGTACAAAAAAGCAGGCT**ATG** <u>CATCACCATCACCATCAC</u>CTGGTGCCGCGCGGCAGC TTTTTAGATGGAATAGAT

Table I-3: Primers used in sequencing. The sequences are presented in the 5' → 3' direction.

Primer name	Sequence
G448	GGGGAAAGAATAGTAGAC
G449	GCCACACAATCATCACCTGCC
T3	ATT AAC CCT CAC TAA AGG G
T7	AAT ACG ACT CAC TAT AGG G
G502	CAATCAAAGGAGATATACCACG
G550	TCGACCTGCAGGCGCGCCGA

Table I-4: Oligodeoxyribonucleotides used in creation of the LTRs for pB2LTR and in construction of the I-PpoI site inserted in pPPOsite.

Oligo name	Sequence	Description
G515	CTCTCTTAAGGTAGC	I-PpoI upper
G517	GCTACCTTAAGAGAG	I-PpoI lower
G569	CTAGTAGTACTGCTAGAGATTTTCCACAGCATG	3' LTR lower
G570	CTGTGGAAAATCTCTAGCAGTACTA	3'LTR upper
G604	CAGTGAATTAGCCCTTCCAGTACTGGTAC	5'LTR lower
G605	CAGTACTGGAAGGGCTAATTCACTGCATG	5'LTR upper
G448	GGGGAAAGAATAGTAGAC	5'newSeq4IP
G449	GCCACACAATCATCACCTGCC	3'NewSeq4IP

APPENDIX II

Table II-1 PCR programs used in the study

| DS2006 (insert PCR)	DPESÄKE (Colony-PCR)	GW2701 GATEWAY-PCR	GW-cIN
1. 94°C 1 min.	1. 95°C 5 min	1. 96°C 1min 30 sec	1. 96°C 1min 30 sec
2. 94°C 30 sec.	2. 95°C 1min	2. 94°C 1min	2. 94°C 1min
3. 50°C 30 sec.	3. 51°C 30 sec.	3. 52,5°C 30 sec	3. 52,5°C 30 sec
4. 72°C 1 min	4. 72°C 1 min 35 s	4. 72°C 2 min	4. 72°C 1 min
5. 25 x cycles 2-4	5. 25 x cycles 2-4	5. 25 x cycles 2-4	5. 25 x cycles 2-4
6. 72°C 5 min	6. 72°C 6 min	6. 72°C 5min	6. 72°C 5min
7. 4°C ∞	7. 4°C ∞	7. 4°C ∞	7. 4°C ∞

APPENDIX III

5   1) GeneRuler™ 100bp DNA Ladder (MBI)

    2) GeneRuler™ 1kb DNA Ladder (MBI)

    3) GeneRuler™ DNA Ladder Mix (MBI)

    4) 1 Kb DNA Ladder (NEB)

10
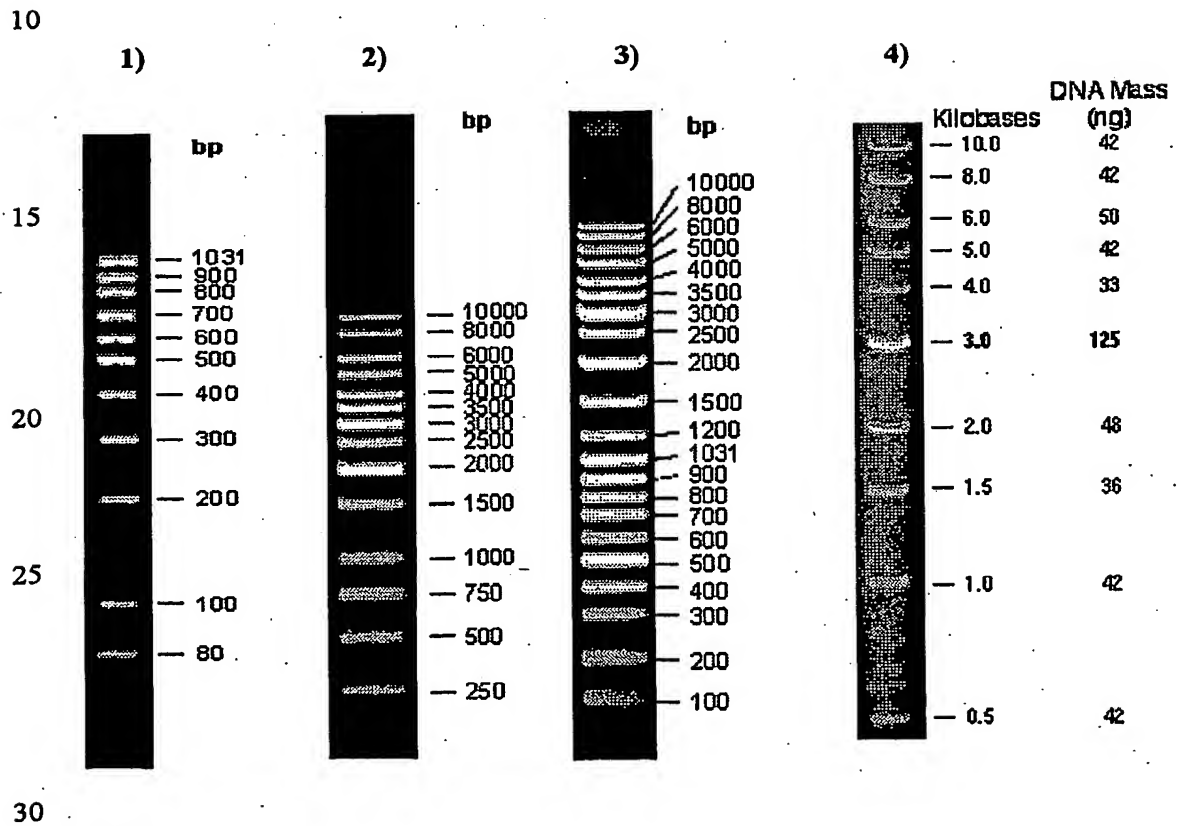
APPENDIX IV:
Buffers and reagents

Ammonium Persulfate 10% (APS)	0,1 g APS, dissolve in 1 ml H_2O Prepare just before use or store at -20°C.
APA-buffer	0,1 M $NaHCO_3$ 1 mM $MgCl_2.6 \cdot H_2O$ Adjust pH with NaOH to 9.8
Kodak (Transfer buffer)	100 ml 10x Kodak stock solution 200 ml methanol 700 ml H_2O 1 L
Kodak (Transfer buffer) 10 x stock solution	30,3 g Tris base 30,3 g Glycine Add H_2O to 1000 ml
LB-solution	10 g Bacto-tryptone 5 g Bacto-yeast extract 10 g NaCl Adjust pH to 7.5 with 1M NaOH, fill to 100 ml with H_2O
PBS-buffer	276 mM NaCl 16 mM Na_2HPO_4 10,7 mM KCl 2,9 mM KH_2PO_4
Ponceau protein staining solution	0,2% Ponceau S 3% TCA
Running buffer 5 X, pH 8.3	15 g Tris base 72 g Glycine 5 g SDS (Sodium dodecyl sulphate) Add H_2O to 700 ml. Adjust pH. Add water to a final volume of 1000 ml. Store at +4°C.
SDS-PAGE Sample buffer (2x)	2,5 ml 0,5M Tris-HCl (pH 6.8) 4,0 ml 10% SDS 2,0 ml glycerol 0,2 ml 0,2% bromophenole blue 1,0 ml β-mercaptoethanol

0,3 ml H_2O

SOC

2 g Bacto-tryptone
0,5 g Yeast-extract
1 ml 1M NaCl
0,25 ml 1 M KCl
1 ml 2 M Mg-stock solution
1 ml 2 M glucose
Adjust volume to 100 ml with H_2O

TAE- buffer

0,04 M Tris-acetate
1 mM EDTA

TBS 0,5M

29,22 g NaCl
3,15 g Tris-HCl
Add H_2O to 1000 ml

TBS- buffer

20 mM Tris-HCl (pH 7.5)
0,5 M NaCl

TBS-Tween-buffer

1 x TBS buffer
0,2% (v/v) Tween20

TE-buffer

10 mM Tris-HCl pH 8
1mM EDTA

Tris-HCl 0,5 M, pH 6.8

6 g Tris base; dissolve in 60 ml H_2O
Adjust pH to 6.8, add H_2O to 100 ml
Store at +4°C.

Tris-HCl 1,5 M, pH 8.8

27,23 g Tris base; dissolve in 80 ml H_2O
Adjust pH to 8.8, add H_2O to 150 ml
Store at +4°C

Preparation of SDS-PAGE gels:

10% Running Gel (thick comb)
6,023 ml Distilled H_2O
3,75   ml 1,5 M Tris-HCl pH 8.8
150   µl 10% SDS-stock
5,00   ml 30% acrylamide/bis-solution (BioRad)
75   µl 10% APS
7,5   µl TEMED

4% Stacking gel (thick comb)
6,10 ml Distilled H_2O
2,50 ml 0,5 M Tris-HCl pH 6.8
100 µl 10% SDS-stock
1,30 ml 30% acrylamide/bis-solution (BioRad)
50 µl 10% APS
10 µl TEMED